UNIVERSITÀ DI PISA
FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
DIPARTIMENTO DI FISICA

TESI DI LAUREA MAGISTRALE

Endogenous and exogenous contributions to market activity: a Hawkes process approach

Relatore:

Prof. FABRIZIO LILLO

Relatore interno:

Prof. RICCARDO MANNELLA

Laureando:

MARCELLO RAMBALDI

ANNO ACCADEMICO 2012/2013

A mio padre

Contents

Introduction

This thesis deals with an interesting problem in the field of mathematical finance and econophysics: the stochastic modelling of the arrival process of market events in continuous time. In this context, a market event is any type of event registered on the market at a definite time, for example the occurrence of a transaction or of a change in the price of an asset. This is a topic of interest both for the academia and for the financial industry, because a faithful description of the time properties of events arrivals is essential for a complete understanding of the market and has important consequences for the construction of trading strategies and for the management of risk.

The research work we present here was conduced in collaboration with the eFX Quantitative Trading group of HSBC in London. They provided us with Foreign Exchange (FOREX) market data and invaluable insight into the market structure and mechanisms.

Market activity, intended as the number of events per unit time, is not a time homogeneous process. Instead, events tend to cluster in time. Its dynamics is influenced both by endogenous and exogenous events. Traders react to other market participants activity and the price movement they generate: intense activity triggers more activity in a self-exciting fashion. Also exogenous shocks play a relevant role, for example when a news is announced, market activity can have a sudden increase. We aim at describing these two mechanisms under the same dynamic model and investigating their relative contribution to market activity.

The mathematical description of this phenomenon comes from the theory of point processes: the sequence of times corresponding to market events is viewed as the realization of a point process on the real line. Hawkes self-exciting point processes are a promising tool to model this phenomenon, since they are characterized by a time clustering structure. These processes were originally introduced in the field of seismology by Hawkes (1971) to describe the aftershocks that follow a major earthquake. They subsequently gained a certain popularity in many other fields including high frequency finance (see Bauwens and Hautsch (2009), Hardiman et al. (2013), Bacry et al. (2013), and Bormetti et al. (2013), among others). Basically, Hawkes processes are Poisson processes with a time dependent intensity. Their self-exciting nature derives from the fact that the intensity depends on the past history of the process and has a jump every time an event is generated, leading to "cascades" of events. The way past events influence the intensity is regulated by the functional form of the so-called *kernel* of the process.

The original research work we present in this thesis consists of three parts. First we analyse in detail the properties of our equity and Foreign Exchange (abbreviated FOREX or FX) datasets (Chapter 3). The equity dataset is relative to trades on the Apple stock at NASDAQ electronic market in August and September 2009. The FOREX dataset is constituted of bid-ask data concerning three currencies pairs, namely EUR/USD, EUR/JPY, USD/JPY, from the interbank trading platform EBS live. The data are relative to the period from 1 January 2012 to 18 December 2012.

In the second part (Chapters 4 and 5), we fit a Hawkes process with different specifications for the kernel on both equity and FOREX data. In this part we focus mainly on the endogenously-generated activity, which, as it turns out, is the main driver of market activity itself. We evaluate goodness-of-fit for different choices of the process kernel. We discuss the strength and limits of the models on both markets and we compare our results with those already published in the literature (Bowsher, 2007; Bacry et al., 2012). We find also that FX market operate close to a critical state, as recently observed in other markets (Hardiman

et al., 2013).

Finally, Chapters 6 and 7 contain two different improvements of the model of the previous chapters. In Chapter 6 we develop a new extension of the Hawkes model that takes into account also the effects of macroeconomic news announcements on FOREX market activity. These kind of announcements are the main exogenous shocks on the FOREX market. In fact, the value of a currency against another one is tightly linked to the market perception of the relative strengths of the respective economies and to the decisions of the respective central banks. Hence, a shift in market perception caused by an unexpected news from the economy can lead to frenetic activity on the market after the announcement. In these cases simple Hawkes models like the ones applied in the second part fail to reproduce the sudden increase in market activity that is observed. On the contrary, our extension captures very well the way news affects market activity, also with regard to the extent the news matched market expectations.

In Chapter 7, we enhance the model presented in Chapter 5 including in the model information on the magnitude of price changes, that was not exploited in the previous models. In particular we employ a kernel dependent on the price changes. We compare the result of this enhanced model with those of the basic model, showing that the new model is able to capture relevant features of the quote change process.

In the first two chapters of the thesis we review some fundamental concepts from econophysics and point process theory that are needed to understand the rest of this work. We give here a brief summary of each chapter:

1. **Structure and dynamics of financial markets** is an introductory chapter. It describes the structure of modern financial markets and introduces some notions and definition from the field of econophysics. It also provides a context for the work presented in this thesis reviewing the debate in the literature over the origin of market activity.

2. **Hawkes processes and their application to finance** starts with a short review of simple point process theory. It then presents two classes of point process that are of interest in finance: the ACD class and the Hawkes class. The latter is described in detail since is of central importance for this thesis work. In particular, the issues of parameter estimation, process simulation, and goodness-of-fit test are discussed. The chapter concludes with a brief survey of the literature of the application of Hawkes processes in the modelling of financial high frequency data.

3. **Data and methods** consists of three parts. The first presents the datasets we used in our empirical tests. We give a description of the original datasets and we explain the filtering procedures we adopted. The second part contains an original analysis of the properties of the financial data we employ. Aspects such as the variations of the level of activity during the day and the distribution of the time intervals between two subsequent market events are examined in some detail. The last part of the chapter is devoted to the illustration of the software and libraries we used in developing the computer code for the various parts of this work.

4. **Modelling equity data with Hawkes processes** presents the results of our first empirical test of a Hawkes model on equity transactions data. We first explain the additional transformation we applied to the dataset. We then detail the model specifications. Finally, we examine the results and discuss the performance of the model.

5. **Modelling FOREX data with Hawkes processes** is analogous to Chapter 4, but now we apply the Hawkes model to FOREX data. With respect to Chapter 4, here we adopt some slight modification of the model. We conclude with an analysis of the merits and limits of the model in this context.

6. **Hawkes process description of exogenous contributions to market activity** represents our main original contribution. We develop an extension of the model presented in the previous chapters that deals with the arrival of news on the market. The aim of this extended model is to reproduce the impact of news arrival on market activity. We start the chapter with an analysis of news impact on FOREX market, also in relation to market expectations towards the announcement. We then

enunciate our model and discuss its main mathematical properties. Next, we report the results of the application of the model and we discuss its performance.

7. In the final chapter, **Expanding the information set: Hawkes processes with kernel dependent on price changes**, we present a modified version of the model of Chapter 5. Now the magnitudes of price changes is included in the model, going in the direction of providing increasingly more complete and faithful models of the market.

We close the thesis summarizing our results and highlighting the main points in the Conclusion.

Chapter 1

Structure and dynamics of financial markets

This is an introductory chapter. Section 1.1, starting from the efficient market hypothesis, reviews some empirical observation regarding financial data and market activity. Then, the debate in the literature on the origins of price changes is briefly summarized. The concepts exposed in this section provide the theoretical background on which this thesis develops. The main reference are Bouchaud and Potters (2003), Chakraborti et al. (2009), Hardiman et al. (2013) and Filimonov and Sornette (2012). Other references are cited in the text.

In Section 1.2 we give an overview of the structure and functioning of modern financial market, with emphasis on those which are investigated in this thesis, namely the *equity* market and the *FOREX* market. We also introduce some definitions and concepts that are used throughout this thesis work. The main references are Bouchaud and Potters (2003) and King et al. (2011), though many textbooks cover these topics.

1.1 The efficient market hypothesis and the origin of price changes

One of the central and most debated ideas in finance is so-called efficient market hypothesis, and a lot of financial literature deals with tests of its predictions. One of these is that price moves are driven mainly by changes in economic fundamentals. However, it has been argued that prices move too much for this being the case and that indeed price moves are driven mostly by the trading activity itself. This debate is the context in which this thesis is inserted. Before exposing the main points of this debate, we review some necessary concepts from the econophysics and quantitative finance literature.

1.1.1 Efficient market hypothesis

One of the recurrent themes in the study of financial market is that of their efficiency. Markets are complex systems. They can be thought of as "engines" through which information is transformed into prices. The classical paradigm in finance is that markets are highly efficient in the determination of the most rational price of the traded asset given the available information. Samuelson (1965) was the first to give a formal definition of market efficiency, and proved that, in an informationally efficient market, price changes must be unforeseeable if they fully incorporate the expectations and information of all market participants. Malkiel (1999) gives the following definition of an efficient market:

"A capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set (...) if security prices would be unaffected by revealing that information to all participants. Moreover,

efficiency with respect to an information set (...) implies that it is impossible to make profit by trading on the basis of that information set."

It is clear from this definition that it is necessary to specify what information set is considered in order to correctly formulate the efficiency hypothesis. Depending on the information set, it is possible to distinguish various degrees of market efficiency (see Roberts (1967)):

- *Weak-form efficiency*: the information set includes only the history of prices or returns themselves;
- *Semistrong-form efficiency*: the information set includes all the publicly available information such as history of prices, analysts reports, corporate earnings and so on.
- *Strong-form efficiency*: the information set includes all information, counting public information as well as any private information known to any single market participant.

The mathematical consequence of the efficiency hypothesis on price is that the price process P_t is a *martingale*[1] with respect to the information set $\mathcal{F}_t$, and in particular:

$$\mathbb{E}\left[P_{t+1}|\mathcal{F}_t\right] = P_t, \tag{1.1}$$

i.e. the best prediction of the future price is the actual observed price.

The efficient market hypothesis is tightly linked to the concept of *arbitrage*. An arbitrage opportunity is present in a market when an economic actor can devise a trading strategy which is able to provide her a financial gain *continuously* and *without risk*. As an example of an arbitrage opportunity suppose that the relative rates between Euro, US Dollar and Yen were EUR/USD$= 1$, USD/JPY$= 10$, and EUR/JPY$= 11$. Then, an investor holding one EUR, would immediately convert it first to 11 JPY, then to 1.1 USD and finally back to 1.1 EUR, making a risk-free profit of 0.1 EUR. In a perfectly informationally efficient market prices should *rationally* reflect the information available, then a case like the one in the example should not even presents, since clearly EUR/JPY is mispriced given the knowledge of EUR/USD and USD/JPY rates. In its strict form, an efficient market is an idealized system. In actual markets, residual inefficiencies are always present. More realistically, we can say that in an efficient market, the exploitation of such arbitrage opportunities leads to their disappearance in a very short time period.

1.1.2 Price, returns and volatility

It is worth noting that there is not, at a certain time t, a unique price for an asset. In fact, as we shall see in Section 1.2.3, there is a price for buyers (*ask*) and another one for sellers (*bid*), and the last traded price is often a different one. The best prices available to buy or sell at a certain time are called the best *quotes*. Usually, the mid price between bid and ask is taken as the reference price at time t. We denote with P_t the price at time t, and where not otherwise stated it is intended to be the mid price.

Very often in finance the focus is not on the price itself but rather on the *return* of an asset. Different definitions of returns exist, however, the most widespread are the *simple return*

$$R_t = \frac{P_{t+1} - P_t}{P_t} = \frac{P_{t+1}}{P_t} - 1, \tag{1.2}$$

and the *log-return*

$$r_t = \ln(1 + R_t) = \ln\left(\frac{P_{t+1}}{P_t}\right). \tag{1.3}$$

The *compound* return over k periods is given respectively by

$$R_t(k) = (1 + R_t) \cdot (1 + R_{t-1}) \cdots (1 + R_{t-k+1}) - 1 = \prod_{i=0}^{k-1}(1 - R_{t-i}) - 1, \tag{1.4}$$

[1] A rigorous definition of martingale will be given in the next chapter. The property (1.1) is sufficient for the purposes of the present chapter.

for the simple returns, and by

$$r_t(k) = r_t + r_{t-1} + \cdots + r_{t-k+1} = \sum_{i=0}^{k-1} r_{t-i} \tag{1.5}$$

for the log-returns.

Returns are calculated over a certain interval of time (a year, a month, a minute, etc.), for this reason a discrete index is used, despite the price is a continuous time process. For example, considering daily returns, r_t refers to the return on day t, r_{t+1} to the return on the next day and so on. The simplicity of the last equation is one of the reasons why log-returns are largely employed. Typically, over short time scales, returns are small numbers, and hence the difference between simple and log returns is almost always negligible. In fact, for $|R_t| \ll 1$, $r_t = \ln(1 + R_t) \approx R_t$.

The rate of diffusion of price is called the *volatility* and it is typically denoted with σ. Unlike price, the volatility is not directly observable and has to be estimated. The volatility is often estimated by calculating the standard deviation of returns on a given time scale:

$$\sigma_T = \sqrt{\mathrm{Var}\,[r_t]}. \tag{1.6}$$

So, for example the volatility of daily returns over one year is estimated by the daily returns standard deviation. Volatility is of central importance in finance. It presents some distinctive properties, such as clustering and long-range autocorrelation, that reveal the non-trivial structure of the return process itself. Moreover, volatility is still often chosen as the measure of risk associated to a given investment. This choice has its roots in the central limit theorem. In fact, we can decompose the log-return over an interval T as the sum of $N = T/\tau$ returns over subperiods of length τ, $r(T) = \sum_{i=0}^{N-1} r_i(\tau)$. In the classical approach it is assumed that the returns are independent random variables, then, in the limit $N \to \infty$, $r(T)$ becomes a Gaussian variable. As such, it is completely characterized by its mean and its variance σ^2. Hence, any measure of risk must be based on σ.

However, as we shall see in the next Subsection, returns are found to be strongly non-Gaussian. Therefore, volatility is not an appropriate measure of risk. Other measures of risk have been developed, for a survey see Bouchaud and Potters (2003).

1.1.3 Stylized facts of returns and volatility

We now briefly outline, following Chakraborti et al. (2009), some of the most remarkable empirical facts regarding returns and volatility.

- **Fat tails.** It is now widely acknowledged that the distribution of returns and log-returns is non-Gaussian and fat-tailed. This has been confirmed by many studies on several markets. Figure 1.1 shows the returns distribution of the S&P 500 stock index for returns calculated over 30 minutes and one day. The distribution is markedly non-Gaussian. It is observed that the return distribution becomes more Gaussian increasing the time horizon over which returns are computed. This phenomenon is sometimes referred to as *aggregational normality*. It must be stressed however, that the convergence is very slow, especially in the tails of the distribution.

- **Short-range return correlation.** The autocorrelation function of the returns, defined as

$$C(l) = \frac{\langle r_t r_{t+l} \rangle}{\langle r_t^2 \rangle} \tag{1.7}$$

  is found to decay to zero exponentially fast after a short time lag. Figure 1.2 highlights this fact. It has been noted that the characteristic time of the exponential decay has reduced significantly over time, in parallel with the diffusion of fully electronics markets. While it was of the order of several

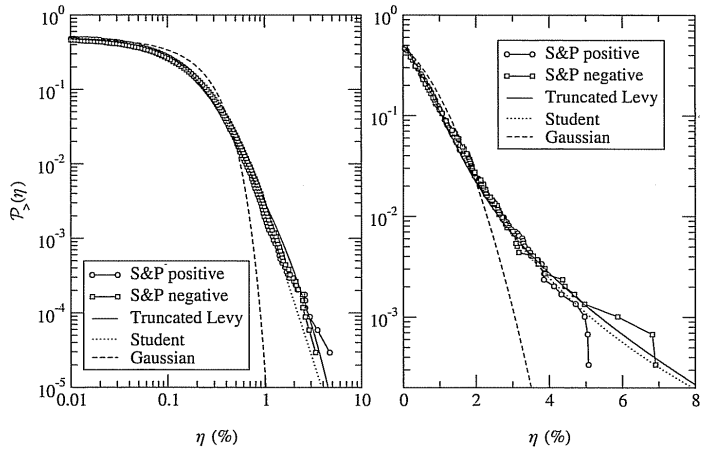

Figure 1.1: Cumulative distribution of positive and negative returns η of the Standard & Poor's 500 stock index over the period 1991-2001. The best fit of some theoretical distributions is also shown. Left: distribution of 30-minutes returns in log-log scale. Right: distribution of daily returns, semi-log plot. Taken from Bouchaud and Potters (2003).

minutes in the Eighties, it is now of the order of one second or even less. The very short-ranged correlation in asset prices and returns is in agreement with the efficient market hypothesis. In fact, past returns are a poor indicator of future returns and hence cannot be used to forecast the future price and make a profit acting accordingly.

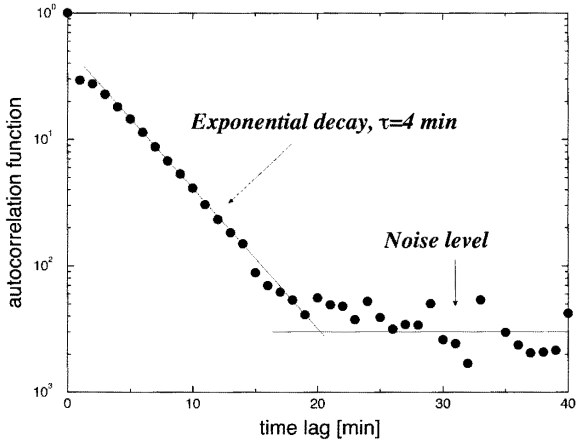

Figure 1.2: Semi-log plot of the autocorrelation function for the S& P 500 returns, sampled at a 1 min time scale. The straight line corresponds to exponential decay with a characteristic decay time of $\tau = 4$ min. It is apparent that after about 20 min the correlations are at the level of noise. Caption and figure taken from Mantegna and Stanley (1999).

- **Volatility clustering** The absence of autocorrelation should not be mistaken for an independence statement. This is very clear if one looks at absolute returns or square returns, where long range slowly-decaying autocorrelation is observed, as in Figure 1.3. This phenomenon is known as *volatility clustering.* and can be summarized with Mandelbrot's words as "large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes" (Mandelbrot, 1963).

These facts are found true also for price changes (rather than returns) and absolute or squared price changes respectively. It is evident already from these few facts that price and return dynamics is a process with a complex correlation structure and a strongly non-Gaussian behaviour.

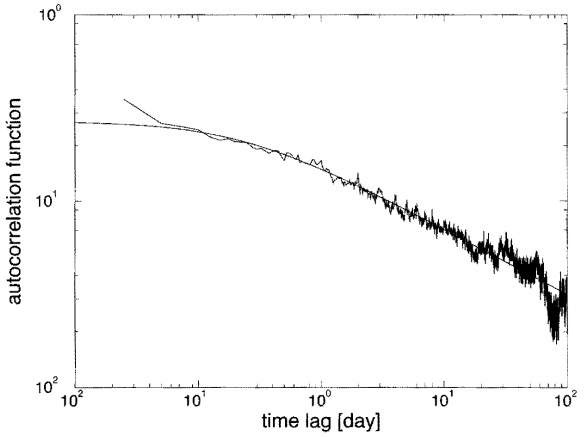

Figure 1.3: Log-log plot of the volatility autocorrelation function calculated from S& P 500 returns sampled at a 1 min time scale. The solid line is a power-law regression fit over the entire range, which gives an estimate of the power-law exponent $\gamma \approx 0.3$ that quantifies the long-range correlations in the autocorrelation function. Caption and figure adapted from Mantegna and Stanley (1999).

1.1.4 Market and trading activity

As the title of this thesis suggests, market activity is one of the central quantities in this work. Although it is intuitively clear that it refers to how much "movement" there is on the market in a certain period, a more precise definition is needed in order to give it a quantitative meaning.

In this thesis with *market activity* we mean the number of best-quote changes in a certain time period. It seems reasonable to take this as a measure of activity, since it is connected to the number of orders submitted or cancelled per time interval and hence to traders' actions and decisions, though we must note that not all orders result in a change of the quotes. We note also that a change in the best quotes almost always implies also a change in the mid price. The exception is when the best bid and the best ask move simultaneously in opposite directions and by the same amount, since in this case the mid price remains unchanged. Market activity is thus also related to how often the price change.

Trading activity is instead defined as the number of transactions that occur in a certain time interval. This quantity is more specific than market activity, since it is linked only to the flow of trades, and it is not as related to price changes as is market activity. In fact, the occurrence of a transaction does not imply that either the bid or the ask must change. Hence, many transactions could happen without changes in the mid price. Moreover, even considering the transaction price as reference price, it is possible to observe many subsequent trades executed at the same price.

From the above definitions, it is clear that the dynamics of market (trading) activity is determined by the quote (trade) arrival process. In the simplest hypothesis that the quote arrival process is a Poisson process, that is, orders arrive in time independently one from each other and with a constant rate λ, then the number of events in an interval τ would follow a Poisson distribution with associated parameter $\lambda\tau$. For $\lambda\tau$ sufficiently large the Poisson distribution is well approximated by a Normal curve with mean and variance $\lambda\tau$.

However, this simple model is disproved by the comparison with empirical evidence. Figure 1.4 shows the inter trade times on the BNP Paribas stock, together with some reference distributions. The exponential distribution, that one would expect if the arrival process was a Poisson process, is clearly inconsistent with the data. The inter arrival times are also found to be strongly correlated (Ivanov et al., 2004). Hence also the independence hypothesis has to be dropped. These properties of inter arrival times cause also market and trading activity to be strongly non Gaussian and autocorrelated. Times of intense activity tend to be followed by high activity. Figure 1.5 shows the cumulative distribution of the number of trades $N_{\Delta t}$ per 15 min of 1000 US stocks, normalized by the respective mean values.

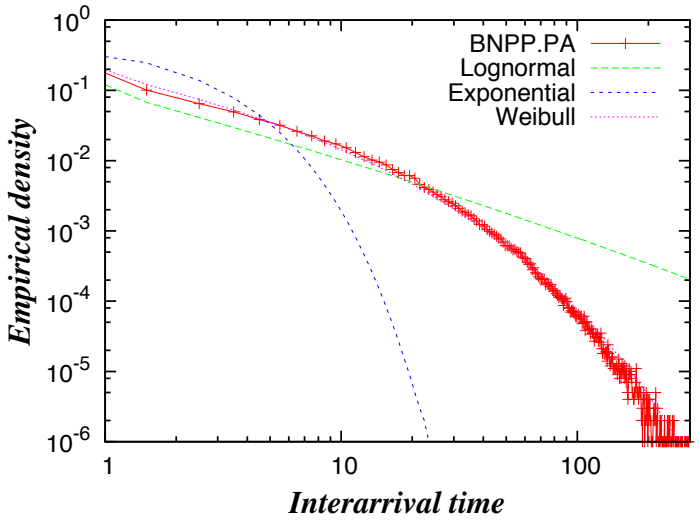

Figure 1.4: Inter trade times distribution for the stock BNPP.PA. From Chakraborti et al. (2009). Axes are logarithmic.

These empirical evidences suggest that a Poisson model for the quote or trade process is totally inadequate. Several alternatives have been proposed such as the ACD and Hawkes models that will be discussed in Chapter 2.

In the next Subsection, a diffusive model of price variation from Plerou et al. (2000) is presented. It will help us evidence the links between the price process and market activity as well as between market activity and volatility.

1.1.5 A diffusion model for price variations

The empirical observation that price time series is indistinguishable from a stochastic process, coupled with the theoretical foundations of price unpredictability provided by the efficient market hypothesis, suggested that the price movements and the motion of a Brownian particle could be described by the same process.

In fact, consider the price change $\Delta P_{\Delta t}$ in a time interval Δt. This is the sum of the $N_{\Delta t}$ individual price changes δp_i that occurred in $[t, t + \Delta t]$:

$$\Delta P_{\Delta t} = \sum_{i=1}^{N_{\Delta t}} \delta p_i. \tag{1.8}$$

This is analogous to the diffusion of a ink particle in water. There, $\Delta P_{\Delta t}$ is the distance covered by the particle, δp_i are the distances covered in between collisions, and $N_{\Delta t}$ is the total number of collisions in the interval Δt. Let $W^2_{\Delta t} = \langle \delta p_i^2 \rangle$ be the variance of the individual steps. Then, for the classic diffusion problem,

1. the distribution $P(N_{\Delta t})$ is a "narrow" Normal, i.e. the standard deviation is much smaller than the mean $\langle N_{\Delta t} \rangle$,
2. the correlation function $\langle N_{\Delta t}(t) N_{\Delta t}(t + \tau) \rangle$ has a short-range exponential decay,
3. also $P(W_{\Delta t})$ is a narrow Normal, and
4. the correlation function $\langle W_{\Delta t}(t) W_{\Delta t}(t + \tau) \rangle$ has a short-range exponential decay too.

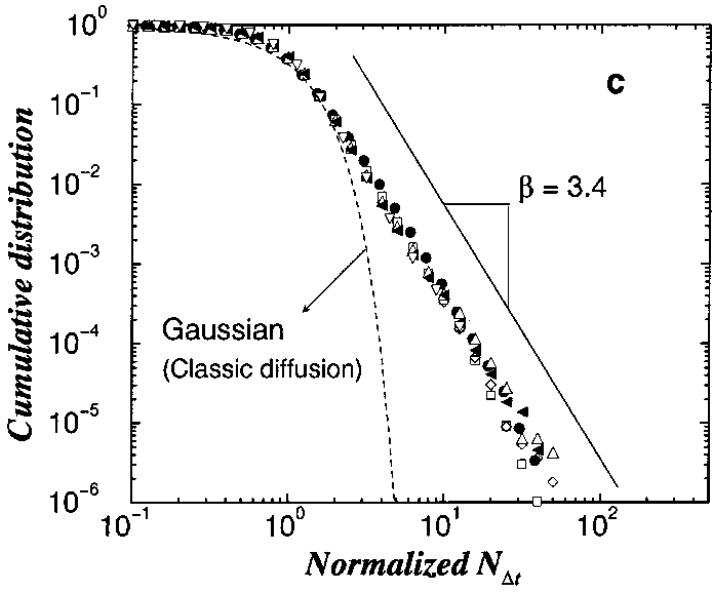

Figure 1.5: Log-log plot of the cumulative distribution of of the number of trades $N_{\Delta t}$ per 15 min of 1000 US stocks. Since each stock has a different average value of $N_{\Delta t}$, we use a normalized number of transactions $n_t = N_{\Delta t}\langle N_{\Delta t}\rangle$. Taken from Plerou et al. (2000).

5. Finally, the random variable $\epsilon = \frac{\Delta P_{\Delta t}}{W_{\Delta t}\sqrt{N_{\Delta t}}}$ is uncorrelated and normally distributed.

These conditions imply that also $\Delta P_{\Delta t}$ is normally distributed and short-range correlated, with a variance $\langle \Delta P_{\Delta t}^2 \rangle = N_{\Delta t} W_{\Delta t}^2 = D\Delta t$, where D is the constant diffusion coefficient. Indeed, the first models of the price process were exactly the same of that for the diffusion of an ink drop.

However, as we have seen, the empirical evidence is that price changes are not normally distributed. Moreover, market activity $N_{\Delta t}$ is not normally distributed either, as is clear from Figure 1.6, nor it is short range correlated. Its distribution presents a power law tail, $P(N_{\Delta t} > x) \sim x^{-\beta}$, and Plerou et al. (2000) found an exponent $\beta = 3.4$ on 1000 stocks analysed. Also $P(W_{\Delta t})$ is found to be strongly non Gaussian and featuring power law tails. Here, Plerou et al. (2000) found the exponent γ that characterize the decay to be $\gamma = 2.9$.

Nevertheless, Plerou et al. (2000) demonstrated that price changes can still be modelled in terms of a diffusion process, albeit an anomalous diffusion one. In particular, the quantity

$$\epsilon = \frac{\Delta P_{\Delta t}}{W_{\Delta t}\sqrt{N_{\Delta t}}} \tag{1.9}$$

is still consistent with a Gaussian-distributed, uncorrelated random variable. The distribution of $\Delta P_{\Delta t}$ is not Gaussian and its properties depend on the characteristic exponent β and γ of the $N_{\Delta t}$ and $W_{\Delta t}$ distributions. In place of the classical diffusion relation $\langle \Delta P_{\Delta t}^2 \rangle = N_{\Delta t} W_{\Delta t}^2 = D\Delta t$, one has $\langle \Delta P_{\Delta t}^2 \rangle = N_{\Delta t} W_{\Delta t}^2 = V_{\Delta t}^2$, where $V_{\Delta t}$ is the local standard deviation of the price changes (analogous to the volatility) and it fluctuates dramatically in time.

In summary, Plerou et al. (2000) found that stock price movements are analogous to a complex variant of classic diffusion, where the analogue of the diffusion constant fluctuates drastically in time. Theirs is a model for price changes and not for returns, however very similar considerations hold also for the return process, and at short time scales the two processes are indeed almost equivalent (Bouchaud and Potters, 2003).

This diffusion model demonstrates the relation between price changes and market activity $N_{\Delta t}$. Furthermore, it also evidences the link between market activity and volatility. Market activity and volatility share similar properties such as power law tails in the distribution and strong autocorrelations. However, this

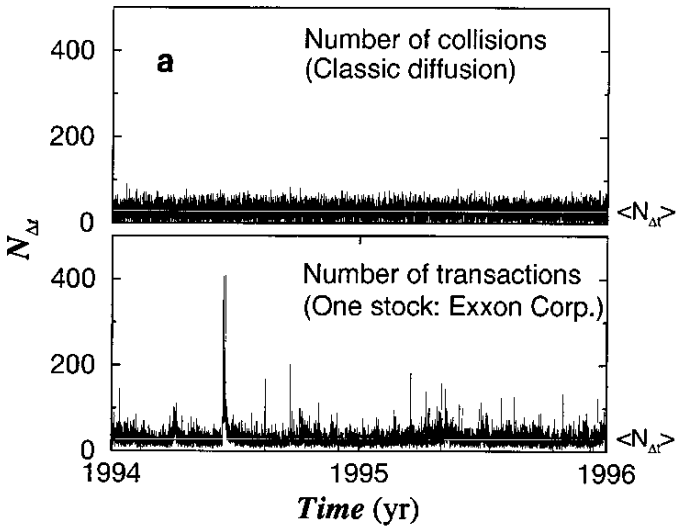

Figure 1.6: The lower panel shows $N_{\Delta t}$ for Exxon Corporation with Δt 15 min. The upper panel shows a sequence of uncorrelated Gaussian random numbers with the same mean $N_{\Delta t} \approx 28$ and standard deviation (≈ 16), which depicts the number of collisions $N_{\Delta t}$ for the classic diffusion problem. Note that in contrast to diffusion, $N_{\Delta t}$ for Exxon shows frequent large events of the magnitude of tens of standard deviations, which have a negligible probability under Gaussian statistics. Figure and caption taken from Plerou et al. (2000).

does not imply that volatility inherits directly its properties from market activity as one could be tempted to conclude (Gillemot et al., 2006).

1.1.6 Do prices move too much?

Until now we have followed mainly an econometric approach: we limited ourselves to the description of empirical data, without caring of the economical implications. Now we turn to the economic interpretation of these empirical facts, and in particular to the debate on the economic reasons for price changes.

In its ideal limit, the efficient market hypothesis states that the market fully absorbs and almost instantaneously reflects in asset prices the information flow (Malkiel, 1999). This means that a movement in the price should correspond to the availability of a new piece of information.

However, since the papers of Shiller (1981) and LeRoy and Porter (1981) the question on whether financial asset prices moves are or not justified by changes in fundamental information has been hotly debated.

In his paper, Shiller (1981) argues that the observed volatility of equities and bonds returns is much higher than the one would expect if prices reflected rational expectations formed on stocks dividends (this debate has been named *volatility puzzle*). Figure 1.7, taken from Shiller (1981) work, highlights this point. The figure shows the detrended actual value of the Dow Jones Industrial Average in the period 1928-1979 confronted with the rational-expectation value inferred ex post from the discounted actual dividends stream. It is possible to note how the first has a much higher variability than the second.

Both papers were highly questioned, and some relevant flaws were pointed out. See LeRoy (2006) for a review on the subject. Nevertheless, many empirical studies have since then confirmed that a large fraction of the volatility of asset prices cannot be explained by changes in their fundamental value (Hardiman et al., 2013; Filimonov and Sornette, 2012). In other words, prices move too much compared with what would be expected from the efficient market hypothesis, even taking into account the costs of gathering information (Filimonov and Sornette, 2012). Moreover, it has been shown (Joulin et al., 2008) that most of the large price jumps observed in the market are not related to the arrival of news in the market place. This is in contrast with the prediction of the efficient market hypothesis that large price moves should

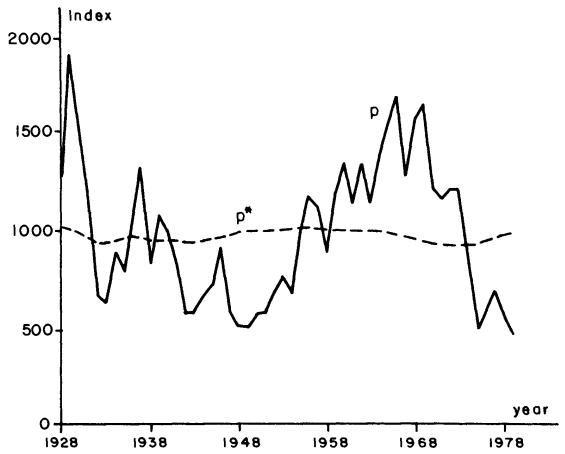

Figure 1.7: Real modified Dow Jones Industrial Average (solid line p) and ex post rational price (dotted line $p*$), 1928-1979, both detrended by dividing by a long-run exponential growth factor. The variable p* is the present value of actual subsequent real discounted dividends, subject to an assumption about the present value in 1979 of dividends thereafter. From Shiller (1981).

only occur in correspondence to the release of news of fundamental economic value.

Many authors (Sornette (2006); Bouchaud (2011) among others) have argued that most market activity is indeed endogenously generated, i.e. it is the trading process itself that triggers more activity even in absence of new information on the fundamental value of the traded asset. In this picture, price dynamics is driven by positive feedback mechanisms involving investors' anticipations that lead to self-fulfilling prophecies. This phenomenon has been called "self-reflexivity" (Soros, 2009).

This debate on the driving forces of market activity has lacked for a long time of appropriate statistical tools to quantify the degree of self-reflexivity. Recently the formalism of Hawkes processes has been applied to try to answer this question (Hardiman et al., 2013; Filimonov and Sornette, 2012).

The work we present in this thesis develops from this theoretical background. Our contribution is towards the development of a faithful econometric model of the quote change process, which is a theme of both theoretical and practical relevance. The theoretical interest resides in the quantification of the endogenous and exogenous contributions to market activity, that is closely related to the debate over the reasons of price changes. On the application side, the dynamics of quote changes is an important block in the development of trading strategies as well as in the management of risk.

Now that we have outlined some important concepts of econophysics and highlighted some reasons why the study and modellization of market activity proves to be an interesting problem, in the next section we give a more detailed and technical description of the financial markets we studied and of their mechanisms.

1.2 Financial markets

Financial markets are marketplaces that allow offer and demand for a certain asset to meet. There exist a variety of different financial products that are traded, including stocks, bonds, currencies and derivatives. In this thesis we will consider only stocks and currencies, though our results could in principle be applied to any asset for which detailed information on trades and quotes dynamic is available. In Subsections 1.2.1 and 1.2.2 we briefly describes the equity market and the FOREX market. Subsections 1.2.3 and 1.2.4 describe the continuous double auction trading mechanism adopted in most electronic markets.

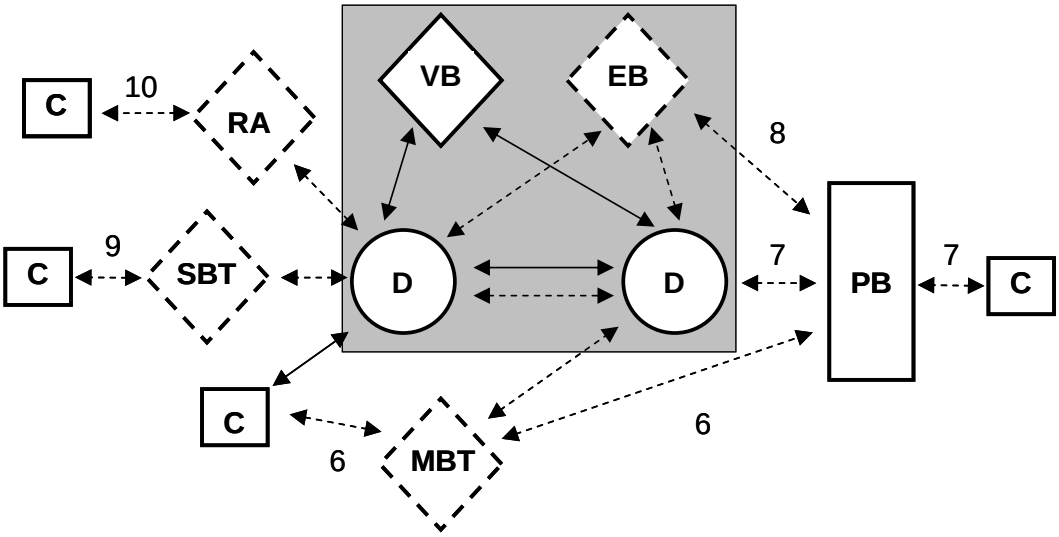

Figure 1.8: Scheme of the structure of modern FOREX market. The D=dealer, C=client, VB=voice broker, EB=electronic broker, PB=prime broker, MBT = multibank trading system, SBT=single-bank trading system, RA=retail aggregator. Solid lines represent voice execution methods. Dashed lines represent electronic execution methods. Taken from King et al. (2011).

1.2.1 Stock markets

The part of a company that belongs to is owners is called *stock*. Often companies have more than a single owner and their stock is broken down into *shares* (or *stocks*). The owners are then called *shareholders* (or *stockholders*). Generally speaking, the terms "stocks" and "shares" are subsumed under the term *equity* (Bouchaud and Potters, 2003). Companies can be either public or private. In a public company, equity can be bought and sold freely in the financial market, while for private companies this is not possible and one has to contact the owners directly.

Public companies shares are traded at *stock exchanges* (or *stock markets*). These are institutions that operate the electronic and physical infrastructures that allow the trading to occur. Example of stock markets are the NASDAQ market in New York and the LSE in London. Nowadays the large majority of stock trading occurs through electronic platforms, most of which are based on the continuous double auction mechanism described in Section 1.2.3.

1.2.2 FOREX market

The FOReign EXchange market (FOREX) is by far the largest marketplace in the world with a daily turnover in excess of US$1 trillion. It is a decentralized global market, meaning that trading occurs through a network of dealers without passing through a regulated exchange as in the case of stocks. Despite its huge size in fact, the FOREX market remains highly unregulated. Its complex structure is sketched in Figure 1.8. Without entering too much into details, the FOREX market structure can be summarized as follows.

- The core of the FOREX market is constituted by big global commercial and investment banks that act as *dealers* (D), i.e. they offer themselves as trading counterpart to clients, they buy if the client wants to sell and vice versa. It is the dealer that sets the prices at which he is willing to trade (maximum price at which he buys, minimum price at which he sells), the client can only accept or refuse them. Clients of these big banks can be smaller banks, other institutional investors (hedge funds, pension funds, big multinational companies, etc.), and also individual investors.

- Major dealers also trade between them to rebalance their positions. They do this either via telephone (VB) or increasingly more via electronic interbank brokers (EB). These are companies that run an electronic trading platform to which all major banks and some hedge funds are connected and where they can anonymously trade one with each other. There are essentially two interbank electronic markets, one run by EBS[2] and the other by Reuters. Since all major players trade on these platforms, they are of central importance in determining the price on all other markets. In fact, the quotes on the interbank markets are used by dealers as starting point to set the prices they offer to clients.

- Clients of big banks can be either end clients or can be themselves smaller dealers. An example of the first group are big multinational companies, that being active in many countries can have the necessity to rebalance their cash reserves from a currency to another. To the second group belong, for example, regional banks that act as clients to the major dealers and as dealers to their own clients (retail investors, small companies, etc.)

In recent years, FOREX market has become much more reliant on electronic trading. Most of the trading happens electronically, though a non-negligible fraction is still conduced by phone. In this thesis we concentrate on the electronic interbank trading platform EBS, which is the most important for EUR/USD, EUR/JPY and USD/JPY. As most electronic markets, it is based on the continuous double auction mechanism described below.

1.2.3 Market mechanisms

As we anticipated most electronic stock markets and also the interbank FOREX market operates through the continuous double auction mechanism. Traders can operate with two types of orders, *limit orders* and *market orders*. A limit order is an order that specifies:

- *Direction*, i.e. if the trader wants to buy or sell.

- *Quantity*, the number of shares (or unit of the asset) the trader is willing to trade (called the *volume*).

- *Acceptable price*, the maximum price at which the trader is willing to buy (*bid* price) or the minimum price at which he is willing to sell (*ask* price).

Limit orders can also be cancelled at any time.

A market order instead specifies only direction and quantity. The price for a market order is the best available on the market at the time the order arrives. All limit orders are gathered in the so-called Limit Order Book. The order book is the list of all buy and sell limit orders and for a given price it shows the total volume available to trade. Figure 1.9 shows how a typical order book looks like at a given instant of time. The *mid price* or *mid quote*, m, is the midpoint between the best bid, b, and the best ask, a: $m = \frac{b+a}{2}$. The difference between the best ask price and the best bid price is called the *bid-ask spread*, this is an important quantity as it gives an order of magnitude of the transaction costs. In fact, considering the mid price as the "fair" price, the cost of a market order is half the spread.

The order book is not a static object, its dynamics is governed by the flow of limit orders and market orders. A set of priority rules determines the sequence in which orders are executed. Price priority is basic. For example, a sell order at 99 will be executed before one priced at 100. The second priority is usually time, limit orders at the same price are executed following the first-in first-out scheme.

[2]The EBS market will be invesitgated in this thesis.

Figure 1.9: Snapshot of the limit order book for the Google stock. Taken from `http://www.ece.cmu.edu`.

1.2.4 Price discreteness

All markets fix a limit to the minimum difference between two prices, and this is called the *tick size*. The ratio between the tick size and the price plays an important role and influences the book dynamics. The fundamental point is whether a change of one tick determine or not a significant price variation. For example, a tick size of 0.01 is small for a stock with an average price of 100, while it is relevant for one priced around 1. In the first case, we speak of small-tick assets. In the second of large-tick assets. Prices of large-tick assets move almost always by one tick, while for small-tick assets the increment distribution is wider.

1.2.5 Liquidity

To conclude this chapter, we concisely review a concept that is often encountered in econophysics and quantitative finance, namely that of the *liquidity* of a market. Here we follow Hasbrouck (2007).

Though a universal and precise definition of liquidity does not exist, some qualities that characterize a liquid market are nevertheless generally accepted. Liquidity is sometimes defined as "depth, breadth, and resiliency":

- Depth: in a liquid market, it is possible to find many participants willing to trade above as well as below the current level of price.
- Breadth: a liquid market has many participants, none of whom has a predominant position.
- Resiliency: price variation associated to any single trade are small and fade quickly away.

Liquid markets typically see high volumes of trading and low costs. It is possible to characterize the agents as suppliers or demanders of liquidity. Suppliers of liquidity are those agents that offer themselves as counterparts, on electronic markets this is done by entering limit orders. This category is traditionally associated with the financial services industry, that is, the brokers, dealers and other intermediaries. The suppliers are "passive agents", they are available for trading but they do not initiate the trade. They are said to "make the market". Demanders, instead, are those that initiate the trade, accepting the proposals of the passive side through market orders. They are the active side that "take the market". Especially

on modern electronic markets, an agent can quickly switch from being a liquidity supplier to being a liquidity demander. Hence, suppliers and demanders of liquidity can no more be identified with specific categories in many electronic markets.

Chapter 2

Hawkes processes and their application to finance

The aim of this chapter is to introduce the reader to the Hawkes processes family and to give some example of their application to the modelling of financial data.

In section 2.1 we recall some general notions from point process theory and we establish some notation. We limit ourselves to what is needed to understand this thesis work, without entering too much into the mathematical details. The main references for this part are Bauwens and Hautsch (2009), Carstensen (2010), Cont and Tankov (2004), and Daley and Vere-Jones (2003).

Section 2.2 gives a short overview of the Autoregressive Conditional Duration (ACD) models, a class of models alternative to the Hawkes one, which is largely employed in finance. Bauwens and Hautsch (2009) is the main reference for this section.

In section 2.3 the class of Hawkes processes is presented and we discuss the related issues of simulation, parameter estimation and goodness-of-fit test for this type of processes. References for this section are Ogata (1978, 1981); Toke (2011a) and Bowsher (2002, 2007).

Finally, in section 2.4, a selected number of works from the literature of the application of Hawkes processes in finance is briefly presented.

2.1 Fundamental concepts

In many fields of research one is often confronted with phenomena that produce events at certain (random) points in time. Examples of such phenomena range from the decay of unstable elements, to the queries to a web server or the occurrences of earthquakes in a precise location. In all cases we do not know neither the total number of events that will occur nor the times at which they will occur. From a mathematical point of view these phenomena constitute a particular type of stochastic process: a (temporal) point process. In the following we give more rigorous definitions.

2.1.1 Point processes

Point processes are stochastic processes with some distinctive characteristics. Here we start giving the definition of a point process on the real line.

Definition 2.1.1. *Lets consider an interval I on the real line, without loss of generality we take $I = [0, T]$, $T \subset \mathbb{R}_0^+$. A simple, non-explosive point process on the interval $[0, T]$ is a sequence $\{t_i\}_{i\in\mathbb{N}}$ of random variables, $t_i \in T$, such that*

I. $\Pr(0 < t_1 < t_2 < ...) = 1$

II. $\Pr(t_i < t_{i+1}, t_i < \infty) = \Pr(t_i < \infty)$ *for* $i \geq 0$

III. $\Pr(\#\, t_i \in T < \infty) = 1, \forall\, T \subset \mathbb{R}$

In words, a point process is an increasing sequence of real-valued random variables. Furthermore, we call it *simple* if the realization are distinct as long as they are finite. That is, no simultaneous events occurs (property (II)). The last property assures that the number of events in every finite interval is always finite, i.e. the point process is non-explosive. In this thesis we will deal only with simple point processes. A well known example in physics is found in the radioactive decay of unstable nuclei. The sequence of times t_i at which a decay is observed from a radioactive sample constitutes a point process,and, in particular, a Poisson process.

The waiting times

$$x_i = t_i - t_{i-1} \tag{2.1}$$

are called the *durations* of the point process $\{t_i\}_{i\in\mathbb{N}}$. It is clear that the starting point t_0 and the durations $\{x_i\}_{\{i=1,2,\ldots\}}$ completely specify the path of the point process $\{t_i\}$. Their knowledge is thus equivalent to that of the times of the realizations of the process.

To every point process it is possible to associate a counting process $N(t)$ defined by:

Definition 2.1.2. *Let* $\{t_i\}_{i\in\mathbb{N}}$ *be a simple point process, the associated counting process* $N(t)$ *is the right-continuous function given by* $N(t) = \sum_{i\geq 1} \mathbb{1}_{t_i \leq t}$. *Where* $\mathbb{1}$ *is the indicator function:* $\mathbb{1}_A = 1$ *if* $t \in A$ *and* $\mathbb{1}_A = 0$ *if* $t \notin A$. *N(t) is itself a stochastic cadlag*[1] *process with piecewise constant trajectories and unit increments.*

$N(t)$ is the total number of events occurred up to time t. Continuing with the radioactive decay example, $N(t)$ is the total number of decay events observed up to time t. An illustration of a simple point process and of its associated counting process is showed in Figure 2.1. Since there is a one-to-one correspondence between a point process and the associated counting process, a point process can be uniquely represented by a counting process. In fact, the notation $N(t)$ is often used to indicate a point process.

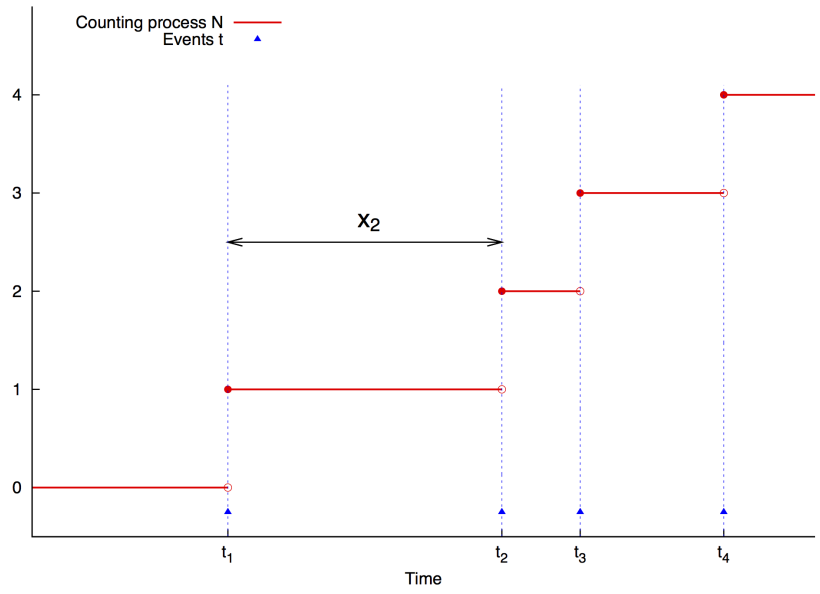

Figure 2.1: Illustration of a simple point process: events, counting process and duration process. Adapted from Toke (2011a).

From Daley and Vere-Jones (2003) a point process is stationary according to the following definition:

[1] A *cadlag* function (from the french "continue à droite, limite à gauche) is function defined on the real numbers (or a subset of them) that is everywhere right-continuous and for which the left limit $f(t-)$ exists $\forall t$.

Definition 2.1.3. *A point process is stationary when for every $r = 1, 2, ...$ and all bounded Borel subsets $A_i, ..., A_r$ of the real line, the joint distribution of*

$$\{N(A_1 + t), ..., N(A_r + t)\}$$

does not depend on t $(-\infty < t < \infty)$.

2.1.2 Marked and multivariate point processes

Often, it is desirable to associate another random variable to each realization of a point process. For example, if we are observing the times at which transactions on a certain asset take place, we can associate to each of them the price paid by the buyer. In this case the process is said a *marked* point process.

Definition 2.1.4. *A marked point process in the interval $T \subset \mathbb{R}$ and with marks in the space $\mathcal{K}$ is a point process $\{(t_i, k_i)\}$ on $T \times \mathcal{K}$ with the additional property that the ground process, i.e. the process without marks, is itself a point process on T.*

We observe that the marks need not to be scalars, they can as well be vector-valued. For instance, with reference to the previous example, we can observe the price paid P together with the amount traded V, the realization of the point process have thus the form (t_i, P_i, V_i).

In many cases it is also interesting to study the interdependence of different point processes, this lead to the concept of *multivariate* point process, which is a special case of a marked point process:

Definition 2.1.5. *A multivariate point process is a marked point process $\{(t_i, W_i)\}$with marks in a finite set $\mathcal{M} = \{1, ..., M\}$.*

In this case the marks are labels of the type of event, for example if we are observing the transactions on two different stocks, say A and B, we have $\mathcal{M} = \{A, B\}$. In the multivariate case we can identify the individual processes as $N^m(t) = \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq t\}} \mathbb{1}_{\{W_i = m\}}$ and the total process as $N(t) = \sum_{m=1}^{M} N^m(t)$. Although a multivariate point process is a special case of a marked point process, there is a clear conceptual difference between the finite-valued marks case and the infinite-valued case that justifies the distinction. In the first case we are interested in the relations between different point processes, in the second case we are dealing with a single component to which we associate a (possibly vector-valued) random variable. Indeed, we can have multivariate point process with associated marks. Continuing with the two stocks example, we can associate to each realization t_i^A or t_j^B the corresponding price paid P_i^A and P_j^B and get a multivariate marked point process.

2.1.3 Filtrations and History

If we interpret the index t of a stochastic process as a time variable, we introduce a dynamic aspect which require us to properly define the notions of information and predictability (Cont and Tankov, 2004). As time passes, more information (e.g the realizations of the process) is revealed. Quantities that are random at time $t = 0$ may become predictable as more information is released. To take into account this dynamic aspect, the concept of *filtration* is introduced:

Definition 2.1.6. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration or information flow on $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family of σ-algebras*[2]

$$(\mathcal{F}_t) \ : \ \forall\, t \geq s \geq 0, \ \mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \tag{2.2}$$

[2] A σ-algebra $\mathcal{F}$ on a space Ω is a family of subset of Ω such that:

- $\Omega \in \mathcal{F}$
- If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
- If $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{F}$ is a numerable sequence of disjoint subsets of Ω, then $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

$\mathcal{F}_t$ is interpreted as the information known at time t, which increase with time. In the case when the past realizations of the process are the only information available, the filtration is called the *natural filtration* or *history* and denoted $\mathcal{H}_t$. Given $\mathcal{F}_t$ it is possible to distinguish quantities that are known from those that are still random at t.

A stochastic process X_t is said to be $\mathcal{F}_t$*-adapted* or *nonanticipating* if its random value is revealed by $\mathcal{F}_t$ (i.e. it is $\mathcal{F}_t$-measurable). A $\mathcal{F}_t$*-predictable* process is a process for which its value at t is known given $\mathcal{F}_{t^-}$, that is, it is $\mathcal{F}_{t^-}$-measurable. A predictable process is also adapted, but the contrary is not true. In particular an adapted left-continuous process is predictable.

2.1.4 Compensator and Conditional Intensity

In martingale-based point process theory, an important concept is that of *compensator*. This, in turn, lead to the introduction of the conditional intensity function, which is of crucial importance in the definition of many point processes including Hawkes processes. In the exposition, we will follow mainly Bauwens and Hautsch (2009). Before introducing these notions we first recall the definition of *martingale*:

Definition 2.1.7. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered probability space with filtration $\mathcal{F}_t$. A stochastic right-continuous process X_t on $[0, T]$ is an $\mathcal{F}_t$-martingale if*

I. X_t *is nonanticipating with respect to* $\mathcal{F}_t$

II. $\forall\, t \in [0, T]$, $\mathbb{E}[X_t] < \infty$

III. $\forall\, s > t$, $\mathbb{E}[X_s | \mathcal{F}_t] = X_t$

If property (III) is substituted by

IV. $\forall\, s > t$, $\mathbb{E}[X_s | \mathcal{F}_t] \geq X_t$

then the process is said a submartingale.

From definition 2.1.7, it is clear that a non-explosive counting process $N(t)$ is a submartingale. As such, it can be decomposed in a zero-mean martingale $M(t)$ and a unique $\mathcal{F}_t$-predictable process, $\Lambda(t)$, called the compensator of $N(t)$, namely

$$N(t) = M(t) + \Lambda(t) \tag{2.3}$$

This decomposition is known as the Doob-Meyer decomposition. Now, since

$$\mathbb{E}[M(t) | \mathcal{F}_t] = 0, \tag{2.4}$$

it follows that

$$\mathbb{E}[N(t) | \mathcal{F}_t] = \mathbb{E}[\Lambda(t) | \mathcal{F}_t]. \tag{2.5}$$

The compensator is then identifiable as the local conditional mean of $N(t)$. Let $\lambda(t)$ be a scalar, positive, $\mathcal{F}_t$-predictable process. Then, $\lambda(t)$ is called the $\mathcal{F}_t$*-conditional intensity* of $N(t)$ if

$$\Lambda(t) = \int_0^t \lambda(u) \mathrm{d}u. \tag{2.6}$$

From (2.5), $\lambda(t)$ can also be defined by the relation

$$\mathbb{E}[N(s) - N(t) | \mathcal{F}_t] = \mathbb{E}\left[\int_t^s \lambda(u) \mathrm{d}u \mid \mathcal{F}_t\right] \tag{2.7}$$

that must hold for all t, s such that $0 \leq t \leq s$. It is possible to obtain a more intuitive interpretation of $\lambda(t)$ letting $s \to t$ in the previous equation

$$\lambda(t^+) = \lim_{s \to t^+} \lambda(s) = \lim_{s \to t^+} \frac{1}{s - t} \mathbb{E}[N(s) - N(t) \mid \mathcal{F}_t] \tag{2.8}$$

Then, $\lambda(t|\mathcal{F}_t)$ is interpreted as the instantaneous expected arrival rate conditional on the available information. In case the point process $N(t)$ is stationary, then $\overline{\lambda} := \mathbb{E}\left[\mathrm{d}N(t)\right]/\mathrm{d}t = \mathbb{E}\left[\lambda(t)\right]$ is constant.

For a multivariate point process we have a conditional intensity function $\lambda^m(t)$ for each subprocess, and the total intensity of the ground process is $\lambda(t) = \sum_{m=1}^{M} \lambda^m(t)$.

2.1.5 Representations of a Point Process

We conclude this section by examining the problem of how to specify the properties of a point process, a question that is of great importance to build point process models of real phenomena. We have seen that to a point process we can uniquely associate a counting process, the process of its durations and a conditional intensity function. Thus, to specify the characteristics of a point process we can set the properties of one of the aforementioned associated processes, this lead to the *counting representation*, the *duration representation* and the *intensity representation* respectively.

The well known Poisson process can be defined following each of the possible representations. In fact, for a Poisson process we have:

Counting representation The Poisson process is the one for which the counting process has independent increments and the number of events in each time interval τ follows a Poisson distribution

$$\Pr\left(N(t+\tau) - N(t) = k\right) = e^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}.$$

Duration representation Let $\{x_i\}_{i\geq 1}$ be independent exponentially distributed random variables with parameter λ, the Poisson process $\{t_n\}$ is then defined by $t_n = \sum_{i=1}^{n} x_i$. That is, a process with independent exponential durations is a Poisson process.

Intensity representation The Poisson process is defined as the point process with constant conditional intensity $\lambda(t) = \lambda$.

The discrete nature of duration models cause them to be unsuitable when the information set has to be updated within a duration spell as is the case for multivariate processes, for example because of an event in other related point process has happened. The ACD class, described in the next section, is an example of process specified via the properties of the durations.

On the opposite side, intensity based models lend themselves perfectly to the multivariate case, as the total intensity is simply the sum of each component intensity and it is easy to make each component depend on the realizations of other-type events. The Hawkes class, which is of central interest in this work, belongs to the models specified via the conditional intensity.

2.2 Duration models: ACD models

Autoregressive Conditional Duration (ACD) models were first introduced by Engle and Russell (1998). This class of models is by far the most used, and a huge literature has been developed on it. Here we will give only a brief overview.

Let x_i be the inter-event duration, possibly standardized by a seasonality function $s(t_i)$, i.e. $x_i := (t_i - t_{i-1})/s(t_i)$. The seasonality function accounts for time-of-day or day-of-week effects (e.g., the openings of related exchanges). In the ACD model, the durations $\{x_i\}_{i=1,\dots,n}$ are modelled in terms of a multiplicative error model:

$$x_i = \Psi_i \epsilon_i, \tag{2.9}$$

where Ψ_i is a function of the past durations and possible covariates, and ϵ_i is a random variable. The main assumption of the ACD model is that ϵ_i is i.i.d. with:

$$\mathbb{E}\left[\epsilon_i\right] = 1, \tag{2.10}$$

so that Ψ_i corresponds to the conditional duration mean:

$$\Psi_i := \mathbb{E}\left[x_i | \mathcal{F}_{t_{i-1}}\right]. \tag{2.11}$$

The conditional duration, Ψ_i, is defined as a function Ψ of the information set $\mathcal{F}_{t_{i-1}}$, it is through this function that the dynamics of the duration process is incorporated. It is common to use a structure of the type:

$$\Psi_i = \Psi(\Psi_{i-1}, ..., \Psi_{i-q}, x_{i-1}, ..., x_{i-p}), \tag{2.12}$$

in which Ψ_i is expressed as a function of the previous q values of Ψ and of the previous p durations x. The simplest ACD model is the one in which a linear parametrization of the conditional duration function is used (*linear ACD model*), limiting ourselves to the case $p = q = 1$, the equation of the linear ACD(1,1) model is:

$$\Psi_i = \omega + \beta\Psi_{i-1} + \alpha x_{i-1}. \tag{2.13}$$

where α, β, and ω are the model parameters. A sufficient condition for Ψ_i to be positive is to ask $\omega > 0$, $\alpha \geq 0$, and $\beta \geq 0$. The process is covariance-stationary provided that:

$$(\alpha + \beta)^2 - \alpha^2\sigma^2 < 1, \tag{2.14}$$

where $\sigma^2 := Var[\epsilon_i] < \infty$.

In linear ACD models it is difficult to allow Ψ_i to depend on function of the covariates without violating the non-negativity restriction. To overcome this difficulty Bauwens and Giot (1997) proposed a class of *logarithmic ACD models*, which has the advantage that no restrictions have to be made on the parameters in order to ensure the positiveness of the process. The parametrization of the logarithmic ACD reads:

$$\ln \Psi_i = \omega + \beta \ln \Psi_{i-1} + \alpha g(\epsilon_{i-1}), \tag{2.15}$$

with:

$$g(\epsilon_{i-1}) = \begin{cases} \ln \epsilon_{i-1}, & \text{log-ACD of type I} \\ \epsilon_{i-1}, & \text{log-ACD of type II} \end{cases} \tag{2.16}$$

where ϵ_i are i.i.d. random variables. The conditions for covariance stationarity are, in this case:

$$\beta < 1, \quad \mathbb{E}\left[\epsilon_i \exp\left\{\alpha g(\epsilon_i)\right\}\right] < \infty, \quad \mathbb{E}\left[\exp\left\{2\alpha g(\epsilon_i)\right\}\right] < \infty. \tag{2.17}$$

Many other specifications for the ACD models exist, see for example Bauwens and Hautsch (2009) and references therein.

2.3 Hawkes processes

Hawkes processes belong to the class of self-exciting processes, characterized by the fact that the intensity is determined by a a weighted function of the time distance to previous points of the process (Bauwens and Hautsch, 2009). A general expression of the intensity function is:

$$\lambda(t) = \varphi\left(\mu(t) + \int_{-\infty}^{t} w(t-s)\mathrm{d}N(s)\right), \tag{2.18}$$

where φ is a generic, possibly non linear function, $\mu(t)$ is a deterministic function of time, and $w(t)$ is a weight function (often referred to as *kernel*). The integral is performed with respect to the counting measure $\mathrm{d}N(t) = \sum_{t_i} \delta(t - t_i)$.

In the model originally developed by Hawkes (1971), φ is a linear function, in particular equation (2.18) take the form

$$\lambda(t) = \mu(t) + \int_{-\infty}^{t} w(t-s)\mathrm{d}N(s) \tag{2.19}$$

This case is the easiest to handle both analytically and computationally, however it requires $\mu(t) > 0$ and $w(t) > 0$ in order to ensure positivity. This in turn excludes the possibility of a self-damping process. In this thesis we will refer always to linear Hawkes processes.

With regard to equation (2.19), taking for simplicity $\mu(t) = \mu$ constant, assuming that the unconditional expectation $\mathbb{E}[\lambda(t)] = \overline{\lambda}$ exists and hence the process is stationary, if we take the expectation on both sides, we get

$$\begin{aligned}
\overline{\lambda} &= \mu + \int_{-\infty}^{t} w(t-s)\mathbb{E}[\mathrm{d}N(s)] = \\
&= \mu + \int_{-\infty}^{t} w(t-s)\mathbb{E}[\lambda(s)]\,\mathrm{d}s = \\
&= \mu + \overline{\lambda}\int_{-\infty}^{t} w(t-s)\mathrm{d}s = \\
&= \mu + \overline{\lambda}\int_{0}^{\infty} w(\tau)\mathrm{d}\tau
\end{aligned} \tag{2.20}$$

If we define $n = \int_0^\infty w(\tau)\mathrm{d}\tau$, for the above equation to have a meaningful solution it is necessary that $n < 1$ and in this case $\overline{\lambda} = \mu/(1-n) \geq \mu$. The condition for stationarity is thus $n < 1$, otherwise, the expression of $\overline{\lambda}$ becomes negative, which is not acceptable and signals that the intensity explodes in finite time. The case $n = 0$ corresponds to a homogeneous Poisson process with intensity μ. The case $n = 1$ is critical, as n approaches 1, $\overline{\lambda}$ diverges. It is possible to preserve stationarity with $n = 1$ if also $\mu = 0$, and certain conditions on the kernel $w(t)$ are satisfied. In this case one has a Hawkes process "without ancestors", as specified by Brémaud and Massoulié (2001). See also Hardiman et al. (2013) and reference therein for more details on this case.

Diverse parametrization exist for $w(t)$, the most common one has been suggested by Hawkes himself (Hawkes, 1971) and is given by:

$$w(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t}, \tag{2.21}$$

where $\alpha_j \geq 0$, $\beta_j > 0$ are the model parameters, the former are scale parameters, while the latter determine the strength of the exponential decay. P denotes the order of the process and has to be selected exogenously. The condition for stationarity is, in this case,

$$\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} < 1. \tag{2.22}$$

Equation (2.21) implies an exponential decay. In application where a long range dependence is needed, the alternative parametrization

$$w(t) = \frac{H}{(t+\kappa)^p} \tag{2.23}$$

which features a power law decay, is often employed. The parameters are in this case H, κ, and $p > 1$.

The multivariate case is obtained from (2.18) by introducing the K-dimensional vector

$$\lambda(t) = (\lambda^1(t), ..., \lambda^K(t))^T$$

each component is thus given by

$$\lambda^k(t) = \varphi\left(\mu^k(t) + \sum_{r=1}^{K}\sum_{t_i^r<t} w_r^k(t - t_i^r)\right). \tag{2.24}$$

Again, one can employ an exponential parametrization for the weight function:

$$w_r^k(t) = \sum_{j=1}^{P} \alpha_{r,j}^k e^{-\beta_{r,j}^k t}. \tag{2.25}$$

$w_r^k(t)$ determine the influence of past r-type events on the k-type intensity. As in the univariate case $\alpha_{r,j}^k \geq 0$ and $\beta_{r,j}^k > 0$.

Hawkes and Oakes (1974) noted that linear self-exciting point processes can be viewed as clusters of Poisson processes. Events are classified into two categories: *immigrants* and *offspring*. Immigrants follow a Poisson process and constitute the centres of so-called Poisson clusters. The cluster representation is then given as follows (taken from Carstensen (2010)):

1. The immigrants follow a Poisson process, N', with intensity function $\mu(t)$, $t \in T$
2. Each immigrant, t_i, generates a cluster, C_i, which consists of events of generations of order $n = 0, 1, ...$ with the following branching structure:
   a) t_i is said to be of generation zero.
   b) Recursively each $t_j \in C_i$ of generation n generates a Poisson process N''_{inj}, of offspring of generation $n+1$ with intensity function $\gamma_j(t) = w(t - t_j)$, $t > t_j$.
   c) The processes N''_{inj} are mutually independent, and given the point that generates them identically distributed modulo the shift.
3. Finally, N consists of the union of all clusters, that is the union of all points for the point process N' and for the N'' point processes.

This description of a Hawkes process in terms of immigrants and offspring paves the way for an interpretation of these processes as composed of "main shocks" (immigrants) and "aftershocks" (offspring) that is interesting both in seismology and high frequency finance. Besides, it gives a sense of the self-exciting character of the process, in that the occurrence of a point "causes" other points to occur. This picture of Hawkes processes qualifies them as branching processes and $n = \int_0^\infty w(t)\mathrm{d}t$ is the branching ratio, i.e. the number of new points that each events triggers on average. Harris (2002) is the standard reference for branching processes. We remark that this interpretation is limited to the linear case, non linear Hawkes processes are not cluster processes.

2.3.1 Simulation

The standard algorithm for the simulation of a Hawkes process is the thinning algorithm developed by Ogata (1981). It is based on the procedure introduced by Lewis and Shedler (1979) for the simulation of non-homogeneous Poisson processes. We first introduce it for the univariate case and then generalize it to the multivariate one. The main references for this part are Ogata (1981); Toke (2011a); Rasmussen (2011); Daley and Vere-Jones (2003).

Consider a point process $N(t)$ on a fixed interval $(0, T]$ with intensity function $\lambda(t)$. Now suppose that it is possible to construct a process with piecewise constant intensity λ^* such that $\lambda(t) \leq \lambda^*$ for every t in $(0, T]$. Finally, let $N^*(t)$ be a locally homogeneous Poisson process with the piecewise constant rate λ^*.

Ogata (1981) proves the following proposition:

Proposition 2.3.1. *Let $t_1^*, t_2^*, ..., t_{N^*(T)}^*$ be the realizations in $(0, T]$ of the process $N^*(t)$. Delete the points t_j^* with probability $1 - \lambda(t_j^*)/\lambda^*(t_j^*)$ for $j = 1, 2, ..., N_T^*$. Then the remaining points $\{t_i\}$ form a point process $N(t)$ with intensity $\lambda(t)$ in the interval $(0, T]$.*

This proposition is the basis for the simulation algorithm. The idea behind the procedure is that when we are at time t, we need to find out where to place the next point $t_i > t$. To this end, we simulate a homogeneous Poisson process on some interval $[t, L(t)]$ for some chosen function $L(t|\mathcal{H}_t)$ that defines the maximum distance we may go forward in time (which may be infinite). Here $\mathcal{H}_t$ is the history of the process up to time t, i.e. $\{..., t_0, t_1, ...t_{n-1}\}$. We choose the rate $M(t)$ of the Poisson process in such a way that:

$$M(t) \geq \sup_{s \in [t, t+L(t)]} \lambda(s) \tag{2.26}$$

so that the previous proposition holds. Once we get the first point t_i from the simulated Poisson process, there are two possibilities: If $t_i > L(t)$, then t_i falls out of the simulation interval and so there are no points in it. If instead $t_i \leq L(t)$, then there may be a point in the interval. The point is accepted, according to Proposition 2.3.1, with probability $\lambda(t_i)/M(t)$. Whether or not the point is accepted, the cycle restart from t_i.

Note that for the procedure to work there must exist the two functions $M(t|\mathcal{H}_t)$ and $L(t|\mathcal{H}_t)$ that satisfy (2.26). For simplicity we write $L(t)$ and $M(t)$ in the following.

The algorithm that implement the procedure described above is the following, n denotes the cumulative number of events in the simulation interval $(0, T]$, and $\mathcal{H}_0$ the history of the process before time 0, i.e. before the simulation starts.

Algorithm 2.3.1. *Ogata's thinning algorithm. Univariate case.*

1. *Set $t = 0$ and $n = 0$. $\mathcal{H} = \mathcal{H}_0$*
2. *Repeat until the termination condition is met:*
   - *Evaluate $M(t)$ and $L(t)$.*
   - *Generate a r.v. s exponentially distributed with parameter $M(t)$ and a r.v. u uniformly distributed in $(0, 1)$*
   - ***If** $s > L(t)$, set $t = t + L(t)$ and return to step 2.*
     ***Else if** $\lambda(t + s)/M(t) > u$, set $t = t + L(t)$ and return to step 2.*
     ***Else** set $n = n + 1$, $t_i = t + s$, $t = t + s$, update $\mathcal{H}$ to $\mathcal{H} \cup t_i$, and return to step 2.*
3. *The output is the list $\{n; t_1, ...t_i\}$.*

This algorithm can simulate various point processes defined by their intensity functions, provided that the conditions for the validity of Proposition 2.3.1 are satisfied. In the case of a Hawkes process, a possible choice for $L(t)$ and $M(t)$ is:

$$L(t) = \infty, \quad M(t) = \lambda(t). \tag{2.27}$$

More precisely, this means that after a point is accepted, say at t_i, the upper bound of the intensity is taken as the value of $\lambda(t_i)$ till a new point is accepted. In fact, in absence of new events the intensity function of a self-exciting process is non-increasing, so the condition (2.26) is satisfied. In Figure 2.2 the result of a simulation for the simplest univariate Hawkes process (exponential kernel of the form (2.21) with $P = 1$) is presented.

For the simulation of a marked point process, only a slight modification of Algorithm 2.3.1 is needed, namely, once the point t_i has been accepted, a mark κ_i is generated from the mark distribution $f(\kappa|t_i)$. Then, (t_i, κ_i) is added to history.

Proposition 2.3.1 is generalized to the multivariate case as follows (from Ogata (1981)):

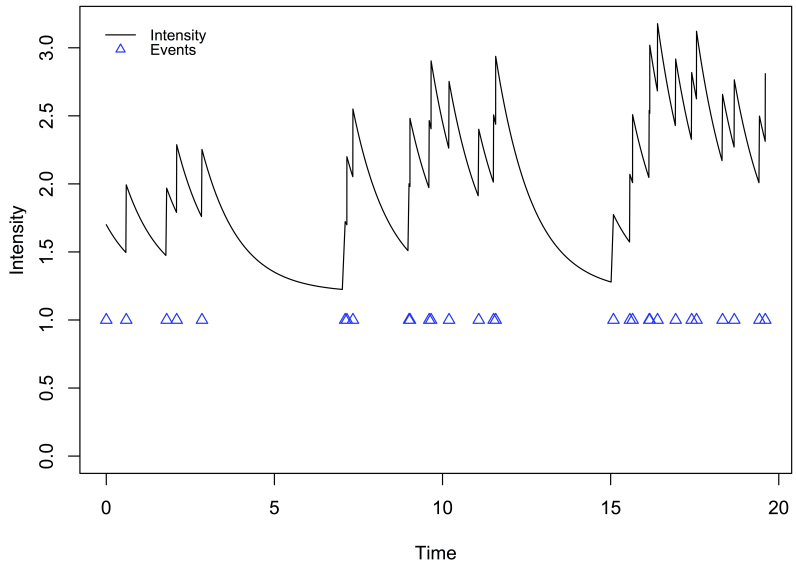

Figure 2.2: Example of a simulated univariate Hawkes process using Ogata's thinning algorithm. The simulation interval is $[0, 20]$. A blue triangle signals the occurrence of an event. A single exponential kernel was employed $(P = 1)$, with $\mu = 1.2$, $\alpha = 0.5$, $\beta = 0.9$. It is possible to note the clustering of the events.

Proposition 2.3.2. *Consider a multivariate point process $N^k(t)$, $k = 1, ..., K$ on a interval $(0, T]$ with joint intensity $\boldsymbol{\lambda}(t) = (\lambda^1(t), ..., \lambda^K(t))$. Suppose that there exists a one-dimensional process λ^* defined pathwise and satisfying*

$$\sum_{k=1}^{K} \lambda^k(t) \leq \lambda^*(t), \qquad 0 < t \leq T \tag{2.28}$$

then set

$$\lambda^0(t) = \lambda^*(t) - \sum_{k=1}^{K} \lambda^k(t). \tag{2.29}$$

Let $t_1^*, t_2^*, ..., t_{N^*(T)}^*$ in $(0, T]$ be the points of the process $N^*(t)$ with intensity $\lambda^*(t)$. For each of the points attach a mark $k = 0, 1, ..., K$ with probability $\lambda^k(t_j^*)/\lambda^*(t_j^*)$. Then the points with marks $k = 1, 2, ..., K$ form a multivariate point process $N^k(t)$ with intensity $\boldsymbol{\lambda}(t)$.

It is convenient to introduce the function

$$I^p(t) = \sum_{k=1}^{p} \lambda^k(t) \tag{2.30}$$

that is the sum of the first p components of the intensity function at time t, furthermore lets take $I^0 = 0$. The thinning algorithm modified for the multivariate case is reported below.

Algorithm 2.3.2. *Ogata's thinning algorithm. Multivariate case.*

1. *Set $t = 0$, $\mathcal{H} = \mathcal{H}_0$*
2. *Repeat until the termination condition is met:*
   - *Evaluate $M(t)$ and $L(t)$.*
   - *Generate a r.v. s exponentially distributed with parameter $M(t)$ and a r.v. u uniformly distributed in $(0, 1]$*

- ***If*** $s > L(t)$, *set* $t = t + L(t)$ *and return to step 2.*
  Else if $I^K(t+s)/M(t) > u$, *set* $t = t + L(t)$ *and return to step 2.*
  Else *set* $n^p = n^p + 1$, $t_i^p = t + s$, *with* p *such that:*

$$\frac{I^{p-1}(t+s)}{M(t)} < u \leq \frac{I^p(t+s)}{M(t)}$$

  . Set $t = t + s$, *update* $\mathcal{H}$ *to* $\mathcal{H} \cup t_i^p$, *and return to step 2.*

3. *The output is the list* $\{\{n^1, ..., n^K; \{t_{i_p}^p\}_{p=1,...,K},\}$

In analogy to the univariate case, a common choice for $M(T)$ in the Hawkes case is the sum of all the components of the intensity function:

$$M(t) = I^K(t) = \sum_{k=1}^{K} \lambda^k(t). \tag{2.31}$$

Again, for the simulation of a multivariate marked point process, once a point of type p, say t_i^p, has been accepted, a mark κ_i^p from the corresponding mark distribution $f^p(\kappa^p|t_i^p)$ is attached to it.

2.3.2 Estimation

For point process models that are specified via the conditional intensity, the likelihood function has a fairly simple expression. This makes maximum likelihood inference a popular choice for these models.

Here we follow the notation of Bowsher (2002). Suppose we are observing a point process in the interval $[0, T]$, we denote $\lambda^*(t) = \lambda(t|\mathcal{H}_{(-\infty,t]})$ the *complete* conditional intensity of the process, that is conditioned on the whole past history. We indicate with $\lambda(t) = \lambda(t|\mathcal{H}_{(0,t]})$ the intensity conditional on the sole information from the observation interval $[0, T]$. Then, from Ogata (1978), the exact log-likelihood on the interval $[0, T]$ for a simple univariate stationary point process with intensity $\lambda(t)$ is

$$\ln \mathcal{L}\left(\boldsymbol{\theta}; N(t)\right) = -\int_0^T (\lambda_{\boldsymbol{\theta}}(s))\mathrm{d}s + \int_0^T \ln \lambda_{\boldsymbol{\theta}}(s)\mathrm{d}N(s), \tag{2.32}$$

where $\boldsymbol{\theta}$ denotes the parameters vector. Ogata (1978) defines also a theoretical log-likelihood under the information from the infinite past

$$\ln \mathcal{L}^*\left(\boldsymbol{\theta}; N(t)\right) = -\int_0^T (\lambda^*_{\boldsymbol{\theta}}(s))\mathrm{d}s + \int_0^T \ln \lambda^*_{\boldsymbol{\theta}}(s)\mathrm{d}N(s). \tag{2.33}$$

Ogata (1978) proves, under some assumptions that are verified for Hawkes processes, some proprieties of the maximum likelihood estimator $\hat{\theta}^T$ obtained by maximizing (2.32) with respect to θ. In particular, $\hat{\theta}^T$ is found to be:

- *Consistent*, i.e. $\hat{\theta}^T$ converges to the true value θ_0 as $T \to \infty$:

$$\forall \epsilon > 0, \quad \lim_{T\to\infty} P\left[|\hat{\theta^T} - \theta_0| > \epsilon\right] = 0 \tag{2.34}$$

- *Asymptotically normal*, i.e.:

$$\sqrt{T}(\hat{\theta}^T - \theta_0) \to \mathcal{N}(0, I^{-1}(\theta_0)) \tag{2.35}$$

  with

$$I^{-1}(\theta) = \mathbb{E}\left[\frac{1}{\lambda^*_{\theta_0}(t)} \frac{\partial \lambda^*_{\theta_0}(t)}{\partial \theta_i} \frac{\partial \lambda^*_{\theta_0}(t)}{\partial \theta_j}\right] = -\mathbb{E}\left[\frac{1}{T} \frac{\partial^2 \ln \mathcal{L}^*_T(\theta_0)}{\partial \theta_i \partial \theta_j}\right] \tag{2.36}$$

- *Asymptotically efficient*, the following asymptotic relation exist between the Hessian of the likelihood function and the information matrix $I(\theta_0)$:

$$-\mathbb{E}\left[\frac{1}{T}\frac{\partial^2 \ln \mathcal{L}_T(\theta_0)}{\partial\theta_i \partial\theta_j}\right] \to I_{ij}(\theta_0) \tag{2.37}$$

  so the variance of the estimator reach asymptotically the lower bound $I(\theta_0)^{-1}$.

These are essentially the standard asymptotic properties of maximum likelihood estimators, however there is an important difference: it is the complete intensity $\lambda^*(t)$ that enters (2.36), and not $\lambda(t)$. Hence is not true in general that $I_{ij}(\theta) = -\mathbb{E}\left[\frac{1}{T}\frac{\partial^2 \ln \mathcal{L}_T(\theta_0)}{\partial\theta_i \partial\theta_j}\right]$, albeit the convergence result (2.37) supports the use of the Hessian matrix $\frac{1}{T}\frac{\partial^2 \ln \mathcal{L}_T(\hat{\theta})}{\partial\theta_i \partial\theta_j}$ to estimate $I_{ij}(\theta_0)$ (Bowsher, 2002).

As for the multivariate case, provided that the components of the intensity $\lambda^m(t)$ have no parameters in common, the log-likelihood can be computed as the sum of the likelihood of each coordinate:

$$\ln \mathcal{L}(\boldsymbol{\theta}; \{t_i\}_{i=1,\dots,n}) = \sum_{m=1}^{K} \ln \mathcal{L}^m(\boldsymbol{\theta}; \{t_i\}), \tag{2.38}$$

where $\ln \mathcal{L}^m(\boldsymbol{\theta}; \{t_i\})$ is defined as in (2.32). Differently from the simple univariate case, however, results on the proprieties of the ML estimators in the multivariate case do not exist at present.

In the particular case of a linear Hawkes process with exponential kernel as in (2.21), denoting with $\Lambda(0,T) = \int_0^T \lambda(s)\mathrm{d}s$ the intensity compensator, the log-likelihood takes the explicit form:

$$\begin{aligned}
\ln \mathcal{L}(\boldsymbol{\theta}; \{t_i\}_{i=1,\dots,N}) &= -\Lambda_{\boldsymbol{\theta}}(0,T) + \sum_{i=1}^{N} \ln \lambda_{\boldsymbol{\theta}}(t_i) \\
&= -\Lambda_{\boldsymbol{\theta}}(0,T) + \sum_{i=1}^{N} \ln \left(\mu(t_i) + \sum_{j=1}^{P}\sum_{k=1}^{i-1} \alpha_j e^{-\beta_j(t_i - t_k)}\right),
\end{aligned} \tag{2.39}$$

where N denotes the number of points in the interval $[0,T]$. Ogata (1981) shows that the log-likelihood function for a linear exponential Hawkes process can be computed using a recursive formula. In fact,

$$\begin{aligned}
R_j(i) &= \sum_{k=1}^{i-1} e^{-\beta_j(t_i - t_k)} \\
&= e^{-\beta_j(t_i - t_{i-1})} \sum_{k=1}^{i-1} e^{-\beta_j(t_{i-1} - t_k)} \\
&= e^{-\beta_j(t_i - t_{i-1})} \left(1 + \sum_{k=1}^{i-2} e^{-\beta_j(t_{i-1} - t_k)}\right) \\
&= e^{-\beta_j(t_i - t_{i-1})} \left(1 + R_j(i-1)\right)
\end{aligned} \tag{2.40}$$

and the log likelihood expression (2.39) can be rewritten as:

$$\ln \mathcal{L}(\{t_i\}_{i=1,\dots,n}) = -\Lambda(0,T) + \sum_{i=1}^{N} \ln \left(\mu(t_i) + \sum_{j=1}^{P} \alpha_j R_j(i)\right) \tag{2.41}$$

with $R_j(1) = 0, \forall j$. This is particularly important for applications since it reduces the time complexity of the likelihood computation from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$, where N denotes sample size.

In the case where the point process is a multivariate Hawkes process, the expression for $\ln \mathcal{L}^m(\boldsymbol{\theta}; \{t_i\})$ is:

$$\ln \mathcal{L}^m(\boldsymbol{\theta}; \{t_i\}) = -\Lambda^m(0,T) + \sum_{i=1}^{N} z_i^m \ln \left(\mu^m(t_i) + \sum_{n=1}^{K} \sum_{j=1}^{P} \sum_{t_k^n < t_i} \alpha_j^{mn} e^{-\beta_j^{mn}(t_i - t_k^n)} \right) \tag{2.42}$$

where:

$$z_i^m = \begin{cases} 1 & \text{if } t_i \text{ is of type } m \\ 0 & \text{otherwise} \end{cases}$$

There exists a recursive formula as in the univariate case, for its definition and for the expression of the likelihood in terms of the recursive function R see Toke (2011a).

2.3.3 Diagnostics

The starting point for the construction of diagnostics for point process models in general is the so-called Random Time Change Theorem. This result allows to transform a wide class of point processes to a homogeneous Poisson process. We report here the statement of the theorem without proof.

Theorem 1. *Let $N^1(t), ..., N^K(t)$ be a multivariate point process formed from event times $t^k_{i=1...n_k}$ with $k = 1, ..., K$. Assume moreover that the intensities $\lambda^k(t)$ are such that $\int_0^\infty \lambda^k(t)dt = \infty$. Define for each k and all $t \geq 0$ the stopping times $\tau_k(t)$ as the solution to*

$$\int_0^{\tau_k(t)} \lambda^k(s)\mathrm{d}s = t \tag{2.43}$$

Then the point processes $\tilde{N}^k(t) = N^k(\tau_k(t))$ are independent Poisson processes with unit intensity. Furthermore the durations of each Poisson process $\tilde{N^k}$ are given by

$$(\tilde{T}^k_{i+1} - \tilde{T}^k_i) = \Lambda^k(t^k_{i+1}, t^k_i) = \int_{t^k_i}^{t^k_{i+1}} \lambda^k(s)\mathrm{d}s \tag{2.44}$$

where the $\Lambda^k(t^k_{i+1}, t^k_i)$ are called the residuals or integrated intensities.

The basic idea then is to use the fact that the residuals $\Lambda^k(t^k_{i+1}, t^k_i)$ are *i.i.d.* **Exp**(1) random variables to construct goodness-of-fit tests. Engle and Russell (1998) proposed a test of excess dispersion (ED) based on the statistic

$$N^k(T)((\hat{\sigma}^2_{\Lambda^k_i} - 1)/\sqrt{8}), \tag{2.45}$$

where $\hat{\sigma}^2_{\Lambda^k_i}$ is the sample variance of the residuals. In the null hypothesis that the residuals are exponentially distributed with unit variance, the statistic is asymptotically distributed as $\mathcal{N}(0,1)$.

Standard Box-Ljung test can be employed to test the independence property of the residuals. Actually it tests the absence of correlations, which is a necessary but not sufficient condition for independence. In addition, goodness-of-fit can be evaluated graphically by means of quantile-quantile plot (QQ-plot)[3] of the residuals against standard exponential distribution quantiles. As a final test, one can compare the durations of the real process and those produced by a simulation of the Hawkes process on the same time interval, using the estimated values of the parameters.

[3] A QQ-plot is a graphical method for comparing two distributions by plotting their quantiles against each other, and it will be used extensively throughout this thesis work. A point (x, y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate). The quantile function $q(p)$ is the inverse of the cumulative distribution function $F(x) = \Pr(X \leq x)$. For a given probability level p its value is the lower $\overline{x}$ such that $F(\overline{x}) = p$. A well known case is for $p = 1/2$, in which $q(p = 1/2) = \overline{x}_{1/2}$ is the median. The QQ-plot is a useful tool to detect differences between two distribution. If the two distribution are equal, the point will lie on the $y = x$ line, any departure from the line highlights discrepancies between the two distributions.

2.4 Examples from the recent literature

In this section we review three selected applications of Hawkes processes to the modelling of financial data. The purpose of this section is to give the reader an idea of the kind of models that have been developed in the literature. The material contained in this section is not strictly necessary for the understanding of the rest of the thesis.

2.4.1 Bivariate Hawkes model on FOREX data

Hewlett (2006) uses a bivariate Hawkes process to model and predict the imbalance of buy and sell trades conditional on history in FOREX market.

Indicating with N_t^{buy} and N_t^{sell} the counting processes for the arrival of buy and sell orders, the author supposes that (N_t^{buy}, N_t^{sell}) is a bivariate Hawkes process with intensity:

$$\lambda_t^{(i)} := \mu^{(i)} + \sum_{j=\{buy,sell\}} \int_{u<t} w_j^i(t-u)\mathrm{d}N_u^{(i)}, \quad i = \{\text{buy, sell}\}\,. \tag{2.46}$$

An exponential parametrization is adopted for the kernel w_{ij},

$$w_{ij}(s) = \sum_{p=1}^{P} \alpha_{ij}^p e^{-\beta_{ij}^p(s)}, \tag{2.47}$$

and the author fixes $P = 1$. The model then reads:

$$\begin{cases} \lambda_t^{\text{buy}} = \mu_{\text{buy}} + \int_{u<t} \alpha_{\text{buybuy}} e^{-\beta_{\text{buybuy}}(t-u)}\mathrm{d}N_u^{\text{buy}} + \int_{u<t} \alpha_{\text{buysell}} e^{-\beta_{\text{buysell}}(t-u)}\mathrm{d}N_u^{\text{sell}} \\ \lambda_t^{\text{sell}} = \mu_{\text{sell}} + \int_{u<t} \alpha_{\text{sellsell}} e^{-\beta_{\text{sellsell}}(t-u)}\mathrm{d}N_u^{\text{sell}} + \int_{u<t} \alpha_{\text{sellbuy}} e^{-\beta_{\text{sellbuy}}(t-u)}\mathrm{d}N_u^{\text{buy}} \end{cases} \tag{2.48}$$

symmetry constraints are imposed, namely, $\mu_{buy} = \mu_{sell} := \mu$, $\alpha_{buybuy} = \alpha_{sellsell} := \alpha_{same}$, $\alpha_{buysell} = \alpha_{sellbuy} := \alpha_{cross}$, $\beta_{buybuy} = \beta_{sellsell} := \beta_{same}$, $\beta_{buysell} = \beta_{sellbuy} := \beta_{cross}$, so that only five parameters are estimated. The estimation is done using maximum likelihood method, subject to non-negativity constraint for α (to ensure the positiveness of the process). The data set is constituted by two months transaction data for EUR/PLN (Euro/Polish Zloty), with resolution of one second. The "buy"/"sell" mark is given according to the direction of the market order which originated the transaction. The data do not feature any record of the volume. Finally, simultaneous trades are aggregated.

The results of the estimation are:

$$\mu = 0.0033s^{-1} \quad \alpha_{same} = 0.0169s^{-1} \quad \alpha_{cross} = 0 \quad \beta_{same} = 0.0286s^{-1}$$

β_{cross} is not estimated since α_{cross} is zero. In Figure 2.3 a QQ-plot of the residuals $\Lambda_n = \int_{t_{n-1}}^{t_n} \lambda_{buy}(t) + \lambda_{sell}(t)\mathrm{d}t$ is reproduced against a standard exponential distribution. Time change theorem states that for intensity based point processes the residuals have a standard exponential distribution, if $\lambda(t)$ is the true intensity of the process. In other words, if the model perfectly describes the observed process then the residuals are exponentially distributed. A Poisson model is shown for comparison. The QQ-plot highlights that the Hawkes model fit is satisfactory, while that of the Poisson is poor.

2.4.2 Modelling microstructure noise

In their paper, Bacry et al. (2013) propose a tick-by-tick price model based on Hawkes processes that aim to reproduce two phenomena observed on high frequency financial data, namely the signature plot

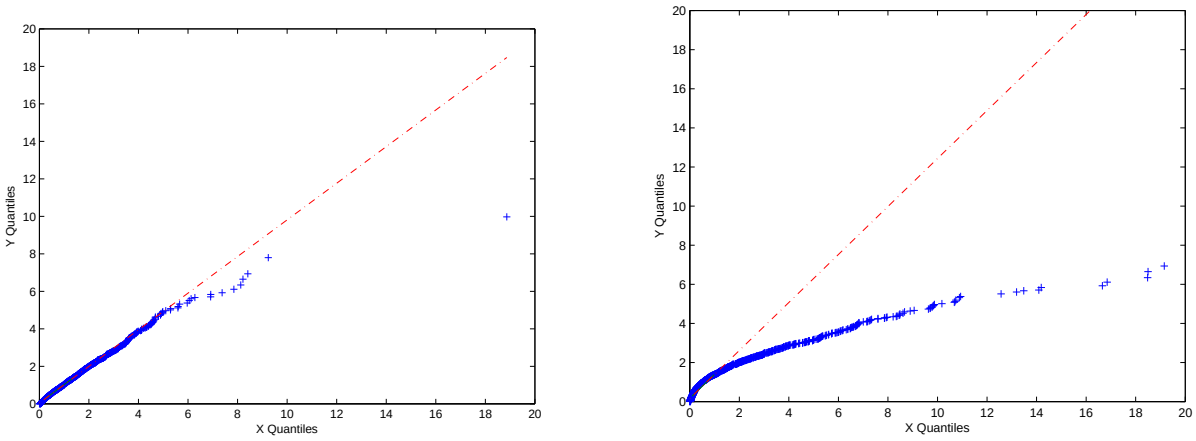

Figure 2.3: Quantile-quantile plots of residuals against exponential for the Hawkes model (left) and a Poisson model (right) on EUR/PLN buy and sell data. For a perfect fit the points should lie on the line $y = x$. From Hewlett (2006)

.

behaviour and the Epps effect. Let $P(t)$ denote the price of an asset at time t, the so-called realized volatility over a time period $[0, T]$ at a scale $\tau > 0$ is defined as

$$\hat{C}(\tau) = \frac{1}{T} \sum_{n=0}^{T/\tau} \left(P\left((n+1)\tau\right) - P\left(n\tau\right) \right)^2 . \tag{2.49}$$

The signature plot is the graph of the realized volatility as a function of τ. The signature plot behaviour refers to the increase in the daily variance $\hat{C}(\tau)$ at small sampling times, i.e. for small values of τ, as visible in Figure 2.4. The realized volatility (2.49) is not stable as τ decreases. This is somewhat counter-intuitive, since one could expect that a higher sampling frequency would increase the precision of the estimation. Instead, this phenomenon make the estimation of the volatility at high sampling frequency unreliable.

The Epps effect denotes instead the decrease of the correlation between two financial assets when the sampling frequency increases. The correlation coefficient estimator $\hat{\rho}(\tau)$ is written as

$$\hat{\rho}(\tau) = \frac{\hat{C}_{12}(\tau)}{\sqrt{\hat{C}_{11}(\tau)\hat{C}_{22}(\tau)}}, \tag{2.50}$$

with

$$\hat{C}_{12}(\tau) = \frac{1}{T} \sum_{n=0}^{T/\tau} \left(P_1\left((n+1)\tau\right) - P_1\left(n\tau\right) \right) \left(P_2\left((n+1)\tau\right) - P_2\left(n\tau\right) \right) \tag{2.51}$$

where $\hat{C}_{11}$ and $\hat{C}_{22}$are the realized volatilities (2.49) of asset 1 and 2 respectively. It is observed that both $\hat{\rho}(\tau)$ and $\hat{C}_{12}(\tau)$ tend to zero when τ decreases.

The authors propose both a univariate model and a bivariate one. In the one dimensional model, the price $P(t)$ of the investigated asset is written:

$$P(t) = N_1(t) - N_2(t),$$

where N_1 and N_2 are Hawkes processes that represent respectively the sum of positive and negative movements of the asset price over the time horizon $[0, T]$, assuming the price changes by units increments. The model is formally similar to that of Hewlett (2006), with intensities given by:

$$\begin{cases} \lambda^1(t) = \mu + \displaystyle\int_{-\infty}^{t} \alpha e^{-\beta(t-s)} \mathrm{d}N_s^2 \\ \lambda^2(t) = \mu + \displaystyle\int_{-\infty}^{t} \alpha e^{-\beta(t-s)} \mathrm{d}N_s^1 \end{cases} \tag{2.52}$$

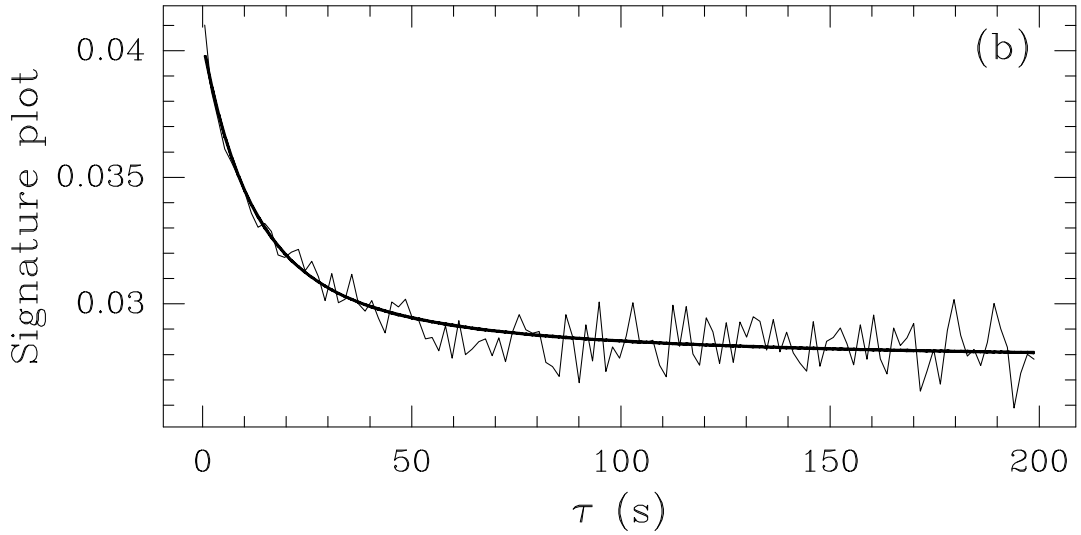

Figure 2.4: Empirical signature plot of the Euro-Bund prices (thin line) and least squares fit of the theoretical 1D model of equation (2.53) (thick line). From Bacry et al. (2013)

.

However, in (2.52) only the cross-excitation terms are present, enforcing the mean reversion empirically observed. Moreover, cross-excitation is set to be symmetric. The authors show that the theoretical signature plot that follows from the univariate model (2.52) is:

$$C(\tau) = \mathbb{E}\left[\hat{C}(\tau)\right] = \Lambda\left(\kappa^2 + (1-\kappa^2)\frac{1-e^{-\gamma\tau}}{\gamma\tau}\right), \tag{2.53}$$

where:

$$\Lambda = \frac{2\mu}{1-\alpha/\beta},\ \kappa = \frac{1}{1+\alpha/\beta},\ \text{and}\ \gamma = \alpha+\beta. \tag{2.54}$$

The theoretical prediction (2.53) fits fairly well to empirical data relative to the Euro-Bund future contract (data from November and December 2009), as can be seen from Figure 2.4. Here equation (2.53) is fitted to the empirical signature plot by least squares method. The parameters of the Hawkes process are then derived using the relations (2.54). As the authors note, using this fitting procedure allows to find the best parameters for reproducing the signature plot behaviour and is therefore appropriate if this is the main focus. However, if the goal is to mimic the arrival times themselves it is better to consider the Maximum Likelihood Estimator, which actually gives similar results for the parameters, but is also less stable with respect to noise and computationally more demanding.

Bacry et al. (2013) also propose a bivariate version of the model, in which two related assets are considered. In analogy with the univariate case the prices of the two stocks are written as:

$$P_1(t) = N_1(t) - N_2(t)$$
$$P_2(t) = N_3(t) - N_4(t)$$

where $\boldsymbol{N} = \{N_i\}_{i=1,\dots,4}$ is a Hawkes process with intensity:

$$\boldsymbol{\lambda} = \boldsymbol{\mu} + \int_0^t \begin{pmatrix} 0 & \varphi_{12} & \varphi_{13} & 0 \\ \varphi_{12} & 0 & 0 & \varphi_{13} \\ \varphi_{31} & 0 & 0 & \varphi_{34} \\ 0 & \varphi_{31} & \varphi_{34} & 0 \end{pmatrix} (t-s)\mathrm{d}\boldsymbol{N_s}, \tag{2.55}$$

where $\varphi_{ij}(t-s) = \alpha_{ij}e^{-\beta_{ij}(t-s)}$. In (2.55), the following assumptions are made:

- self-exciting terms are neglected, $\varphi_{ii} = 0,\ \forall i$;

- upward and downward effects are symmetric within the processes P_1 and P_2, $\varphi_{12} = \varphi_{21}$, $\varphi_{34} = \varphi_{43}$;
- P_1 and P_2 influence each other only in a positive way, $\varphi_{14} = \varphi_{41} = \varphi_{23} = \varphi_{32} = 0$.

Moreover, the authors suppose that $\mu_1 = \mu_2$ and $\mu_3 = \mu_4$, in accordance with the univariate case.

An analytical expression for the covariance matrix and the correlation coefficient is then derived. For the details of the derivation and the complete expression see (Bacry et al., 2013). In Figure 2.5, the empirical correlation coefficient is plotted along with a least squares fit according to the 2D Hawkes model. The agreement is satisfactory given the simplicity of the model, besides, the model is in accordance with the Epps effect. As a final remark, Bacry et al. also discuss the large scale behaviour of their model, and show that is possible to obtain a limiting process which is a multivariate Brownian motion.

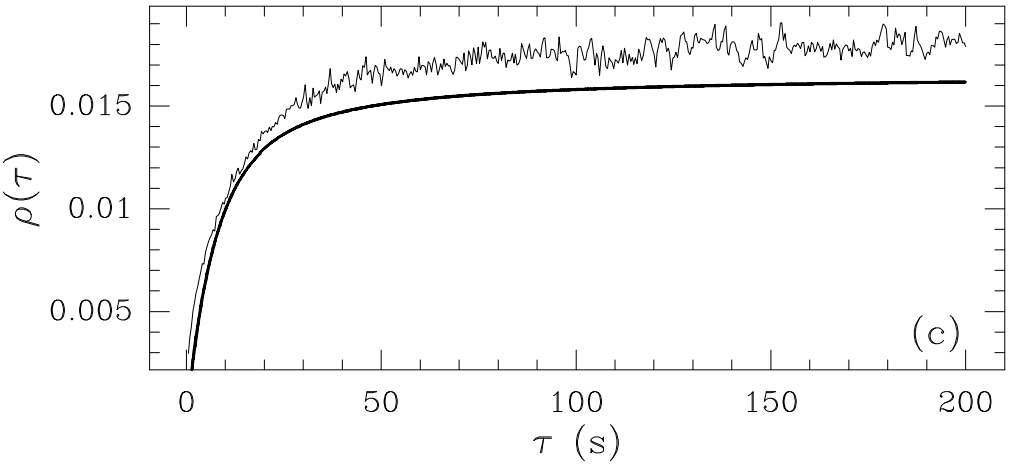

Figure 2.5: Empirical correlation between Euro-Bund and Euro-Bobl prices (thin line) and least squares fit according to the theoretical bivariate model (thick line). From Bacry et al. (2013)

.

2.4.3 Basic order book simulator

Toke (2011b) proposes a basic order book simulator with arrival times of limit and market orders following mutually exciting Hawkes processes. He starts from empirical observations, verified on several markets, that point to the existence of a sort of "market-making" behaviour. What is observed is that, after the occurrence of a market order, a limit order is likely to be submitted more quickly than in the absence of the market order. Otherwise said, according to the author, there seems to be a reaction from the market participants to balance the liquidity removal caused by the execution of the market order. The author also notes that the opposite effect, i.e. increased probability of arrival of market orders after limit orders, is not observed.

The order book model is build starting from a zero-intelligence agent-based market simulator of the kind of the one introduced by Mike and Farmer (2008), and features two agents. The first is a *liquidity provider*, whose characteristics are determined by the following assumptions:

1. submission times of new limit orders are distributed according to a homogeneous Poisson process N^L with intensity λ^L;
2. submission times of cancellations of orders are distributed according to homogeneous Poisson process N^C with intensity λ^C;
3. placement of a new limit order is centred around the same side best quote and is distributed according to a Student's distribution with ν_1^P degrees of freedom, shift parameter m_1^P, and scale parameter $s_!^P$;

4. new limit orders' volume follows an exponential distribution with mean m_1^V;
5. in case of a cancellation, the agent deletes his own order with probability δ.

The second agent is a *liquidity taker*, he only submits market orders, and his behaviour is determined by the below hypothesis:

6. submission times of new market orders are distributed according to a homogeneous Poisson process N^M with intensity μ;
7. market orders' volume follows an exponential law with mean m_2^V.

Finally, agents submit orders on the bid or ask side with probability 0.5. This simple model is referred to by the author as "Homogeneous Poisson" (HP). The model is then modified to include dependence between order flows, in particular, the flow of market and limit orders are modelled by Hawkes processes N^L and N^M, with stochastic intensities λ and μ governed by:

$$\begin{cases} \mu^M(t) = \mu_0^M + \displaystyle\int_0^t \alpha_{MM} e^{-\beta_{MM}(t-s)} \mathrm{d}N_s^M \\ \lambda^L(t) = \lambda_0^L + \displaystyle\int_0^t \alpha_{LM} e^{-\beta_{LMt-s)}} \mathrm{d}N_s^M + \int_0^t \alpha_{LL} e^{-\beta_{LL}(t-s)} \mathrm{d}N_s^L \end{cases} \tag{2.56}$$

Three mechanisms are present here, two are self-exciting ones, "MM" and "LL", the third, "LM", reproduces the "market making" effect mentioned before. Note that the term ML is not included since it is not relevant according to empirical observations. It is possible to switch on and off each term to compare the results of different models.

The parameters are estimated by maximum likelihood method. The data set is constituted by order book data from September 10th, 2009 to September 30th, 2009 (i.e. 15 days of trading) relative to three major stocks quoted on the Paris Bourse (BNPP.PA, PEUP.PA, LAGA.PA) and two futures contracts (FEIZ9, FFIZ9). For each trading day, only 4 hours of data, precisely from 9:30 am to 1:30 pm, are employed. This is done to avoid the opening of American markets and the relative increase of activity.

Once the parameters are estimated, long runs of simulations are performed (24 hours of continuous trading each time). Then, the empirical distributions of market and limit orders is computed and compared with the one observed in real data. Figures 2.6, 2.7, and 2.8 show the results obtained with the different models.

The MM+LL+LM model is the closest to the empirical data, not only with regard to the duration distribution, but also for the spread distributions. In particular, the LM term helps to reproduce the spread distribution.

In conclusion, the work of Toke demonstrates the promising use of Hawkes processes to reproduce the order book dynamics, it also highlights the importance of the cross term LM for a realistic simulation. Limits include very short characteristic times produced by the exponential kernel, as opposed to slow power law decays in real data, and absence of the link with volumes.

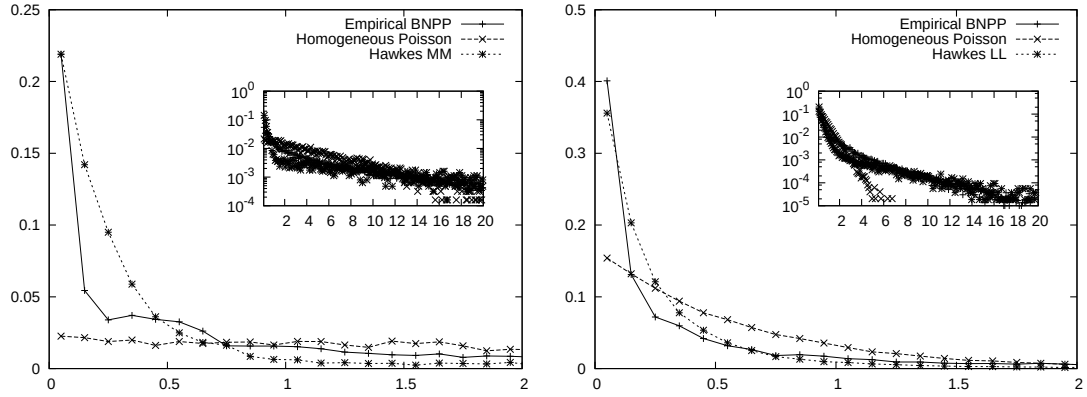

Figure 2.6: Empirical probability function of the durations of market orders (left) and limit orders (right) for three simulations, namely HP, MM, LL, compared to empirical measures. In inset, same data using a semi-log scale. From Toke (2011b)

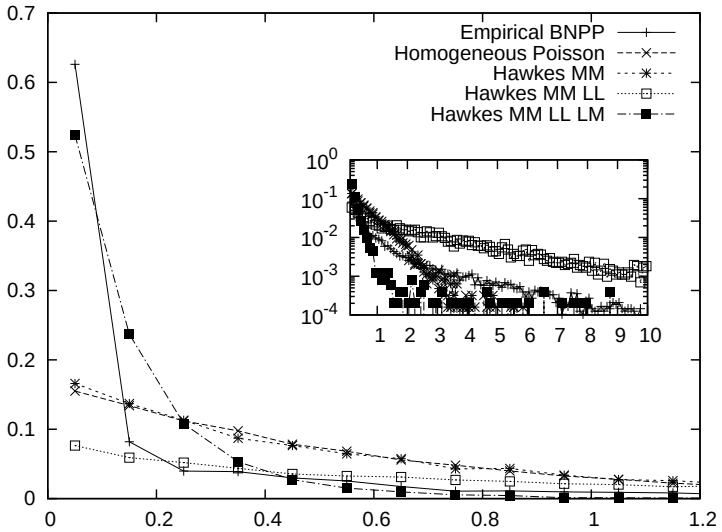

Figure 2.7: Empirical probability function of the time intervals between a market order and the following limit order for three simulations, namely HP, MM+LL, MM+LL+LM, compared to empirical measures. In inset, same data using a semi-log scale. From Toke (2011b)

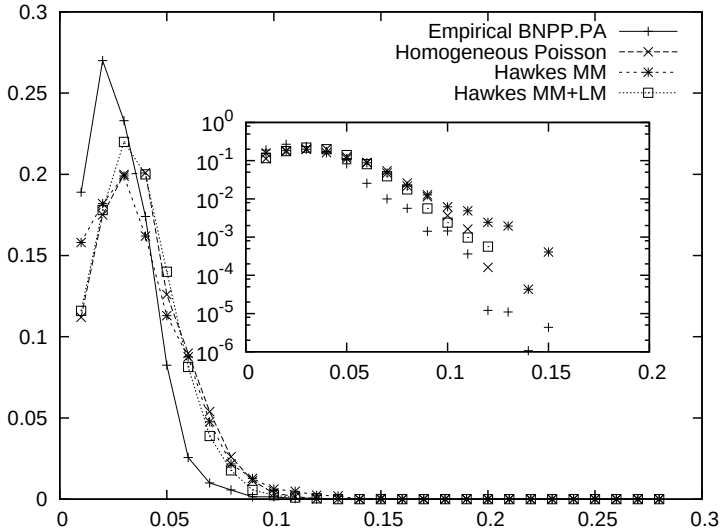

Figure 2.8: Empirical probability function of the bid-ask spread for three simulations, namely HP, MM, MM+LM, compared to empirical measures. In inset, same data using a semi-log scale. X-axis is scaled in Euro (1 tick is 0.01 Euro). From Toke (2011b)

Chapter 3

Data and methods

In this chapter we describe the data on which empirical tests are made and the software and libraries we employed. In section 3.1 we present the financial datasets we use in this work and for each of them the general filtering procedures that were adopted. Section 3.2 presents some descriptive properties of the data in each dataset, all the material in this section is original. Finally, Section 3.3 concerns the programming languages, software and libraries adopted in this thesis.

3.1 Datasets

In this work we have used two different financial datasets. The first reports transactions relative to the Apple Inc. stock registered on NASDAQ between August and September 2009. This dataset was employed in preliminary applications, to check the code and the algorithms and to assess the performance of a simple linear Hawkes model on real financial data with different functional forms for the kernel w.

The second dataset is our main dataset. Most of the work presented in this thesis was performed on this dataset. It consists of data from the interbank FOREX electronic trading platform *EBS live*. Data range from January 1st 2012 to December 18th 2012, for three pairs of currencies, namely EUR/USD, EUR/JPY and USD/JPY. These data were provided by the eFX Quantitative Trading division of HSBC plc. in the context of a research collaboration between HSBC and the Scuola Normale Superiore.

A third dataset contains the list of all the macroeconomic data releases of 2012 and was used for the work that will be presented in Chapter 6.

3.1.1 Equity dataset

These data contain information on the transactions of the Apple stock occurred on the NASDAQ electronic trading platform. The trades in our dataset were registered between 1st August and 30th September 2009, for a total of 42 days of trading, during Regular Trading Hours. NASDAQ opens at 09:30 EST an closes at 16:00 EST.

Time, price and volume information are reported for the trades. The time resolution is one millisecond and all the transactions that happen within the same millisecond are aggregated into a single event and hence are undistinguishable. A sample of the data is reproduced in Figure 3.1. The data had been used in a previous work and they were already preprocessed.

Each line contains the time at which the transaction was registered expressed in milliseconds from 00:00 EST of each day, along with information on the price at which the transaction occurred and on the corresponding volume. The volume, i.e. the number of shares traded, is given with sign. A positive sign means that the transaction was buyer-initiated, while a negative sign stands for seller-initiated transactions. This dataset contains both regular and hidden transactions. Hidden transactions are those that happens outside the order book, since they result from so called hidden orders, a type of order that is not disclosed

day	time	priceNH	volume	priceH	volumeH	mid
1	34219417	14373.00	-214	14372.0	-86	14373.5
1	34219430	14370.00	-100	14372.0	-314	14373.5
1	34219593	0.00	0	14370.7	100	14368.5
1	34219718	0.00	0	14370.0	100	14368.5
1	34220530	14370.00	-200	14369.0	-100	14372.0
1	34220645	0.00	0	14369.0	-100	14372.0
1	34220678	0.00	0	14369.0	-150	14372.0
1	34220705	0.00	0	14368.0	100	14363.0
1	34220711	0.00	0	14369.0	100	14363.5
1	34221500	14366.76	-132	14365.0	-68	14367.5
1	34222096	0.00	0	14370.0	100	14367.5
1	34222523	0.00	0	14369.0	100	14367.0
1	34222525	14369.00	100	14369.5	200	14367.0

Figure 3.1: Sample extracted from the original Apple dataset. The first column indicates the day within the sample (1-42), the second column contains the time of the trades expressed in millisecond from 00:00 of the same day. Third column, if non-zero, is the volume-weighted price for the regular trade. Fourth column is the signed volume of the regular trade. Columns 5 and 6 are the same of 3 and 4 but for hidden orders. The last column is the last volume-weighted mid-price before the trade.

to other market participants and remain invisible in the book till another order of opposite direction matches it. In our study we are interested only in regular transactions.

In order to use them for our purposes, the original data were filtered and processed according to the following procedure:

1. We removed all the lines relative to hidden transaction;
2. We retained only the times column;
3. For each day we set the time origin at 09:30 EST, i.e. at market opening.

At the end of this procedure we have a set of 42 sequences of times, one for each trading day in the original sample.

3.1.2 FOREX dataset

Our data for the FOREX market are constituted by EBS live data from January 1st 2012 to December 18th 2012, a total of 353 days, for three pairs of currencies, namely EUR/USD, EUR/JPY and USD/JPY. Data for each pair are organized in different files, one per day. Every file is a data frame each row of which corresponds to a snapshot of the order book for that pair. The snapshots are taken *every* 100 ms, in particular:

- if a snapshot was taken at time t the next will not be taken until $t + 100\,\text{ms}$[1],
- if between t and $t+\Delta$ the order book does not change, no snapshot is recorded, even if $\Delta > 100\,\text{ms}$.

Each snapshot contains a detailed picture of the order book and a total of 73 column is given. We will limit our description to the most relevant of them. The relevant information to our purposes can be divided into three groups:

- **Time information** For each row, the date plus six different times are given. Their meaning is specified in Table3.1. All times are UTC[2] times.

[1] Actually this does not seem to be strictly true. The presence of intervals of less then 100 ms is discussed later in the text.

[2] Coordinated Universal Time. It is the successor of Greenwich Mean Time (GMT), and for must purposes the two are synonyms.

date	sym	time	ebsReferenceTime	ebsMarketUpdateTime	feedHandlerReceiveTime	feedHandlerPublishTime	eventCaptureTime	bid	ask	paid	given
2012-01-27	EURJPY	09:37:09.106	2012-01-27 09:37:09.100	2012-01-27 09:37:09.104	2012-01-27 09:37:09.105	2012-01-27 09:37:09.105	2012-01-27 09:37:09.106	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:09.309	2012-01-27 09:37:09.300	2012-01-27 09:37:09.304	2012-01-27 09:37:09.308	2012-01-27 09:37:09.308	2012-01-27 09:37:09.309	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:09.411	2012-01-27 09:37:09.400	2012-01-27 09:37:09.405	2012-01-27 09:37:09.409	2012-01-27 09:37:09.410	2012-01-27 09:37:09.411	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:09.511	2012-01-27 09:37:09.500	2012-01-27 09:37:09.505	2012-01-27 09:37:09.510	2012-01-27 09:37:09.510	2012-01-27 09:37:09.511	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:09.609	2012-01-27 09:37:09.600	2012-01-27 09:37:09.604	2012-01-27 09:37:09.608	2012-01-27 09:37:09.608	2012-01-27 09:37:09.609	NA	100.856	NA	NA
2012-01-27	EURJPY	09:37:09.709	2012-01-27 09:37:09.700	2012-01-27 09:37:09.704	2012-01-27 09:37:09.708	2012-01-27 09:37:09.708	2012-01-27 09:37:09.709	100.839	NA	NA	NA
2012-01-27	EURJPY	09:37:09.813	2012-01-27 09:37:09.800	2012-01-27 09:37:09.806	2012-01-27 09:37:09.811	2012-01-27 09:37:09.811	2012-01-27 09:37:09.813	100.840	100.852	NA	NA
2012-01-27	EURJPY	09:37:09.912	2012-01-27 09:37:09.900	2012-01-27 09:37:09.906	2012-01-27 09:37:09.911	2012-01-27 09:37:09.911	2012-01-27 09:37:09.912	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:10.012	2012-01-27 09:37:10.000	2012-01-27 09:37:10.005	2012-01-27 09:37:10.010	2012-01-27 09:37:10.011	2012-01-27 09:37:10.012	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:10.113	2012-01-27 09:37:10.100	2012-01-27 09:37:10.106	2012-01-27 09:37:10.111	2012-01-27 09:37:10.111	2012-01-27 09:37:10.112	100.841	100.856	NA	NA
2012-01-27	EURJPY	09:37:10.213	2012-01-27 09:37:10.200	2012-01-27 09:37:10.205	2012-01-27 09:37:10.210	2012-01-27 09:37:10.211	2012-01-27 09:37:10.212	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:10.310	2012-01-27 09:37:10.300	2012-01-27 09:37:10.305	2012-01-27 09:37:10.309	2012-01-27 09:37:10.309	2012-01-27 09:37:10.310	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:10.415	2012-01-27 09:37:10.400	2012-01-27 09:37:10.406	2012-01-27 09:37:10.411	2012-01-27 09:37:10.412	2012-01-27 09:37:10.415	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:10.514	2012-01-27 09:37:10.500	2012-01-27 09:37:10.506	2012-01-27 09:37:10.511	2012-01-27 09:37:10.511	2012-01-27 09:37:10.513	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:10.613	2012-01-27 09:37:10.600	2012-01-27 09:37:10.606	2012-01-27 09:37:10.611	2012-01-27 09:37:10.612	2012-01-27 09:37:10.613	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:10.709	2012-01-27 09:37:10.700	2012-01-27 09:37:10.704	2012-01-27 09:37:10.708	2012-01-27 09:37:10.708	2012-01-27 09:37:10.709	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:11.110	2012-01-27 09:37:11.100	2012-01-27 09:37:11.104	2012-01-27 09:37:11.108	2012-01-27 09:37:11.108	2012-01-27 09:37:11.109	NA	100.863	NA	NA
2012-01-27	EURJPY	09:37:11.208	2012-01-27 09:37:11.200	2012-01-27 09:37:11.204	2012-01-27 09:37:11.207	2012-01-27 09:37:11.207	2012-01-27 09:37:11.208	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:11.309	2012-01-27 09:37:11.300	2012-01-27 09:37:11.304	2012-01-27 09:37:11.307	2012-01-27 09:37:11.308	2012-01-27 09:37:11.309	NA	100.862	NA	NA
2012-01-27	EURJPY	09:37:11.411	2012-01-27 09:37:11.400	2012-01-27 09:37:11.405	2012-01-27 09:37:11.408	2012-01-27 09:37:11.409	2012-01-27 09:37:11.410	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:12.010	2012-01-27 09:37:12.000	2012-01-27 09:37:12.005	2012-01-27 09:37:12.009	2012-01-27 09:37:12.009	2012-01-27 09:37:12.010	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:12.625	2012-01-27 09:37:12.600	2012-01-27 09:37:12.608	2012-01-27 09:37:12.622	2012-01-27 09:37:12.623	2012-01-27 09:37:12.624	100.840	100.860	NA	NA
2012-01-27	EURJPY	09:37:12.718	2012-01-27 09:37:12.700	2012-01-27 09:37:12.707	2012-01-27 09:37:12.714	2012-01-27 09:37:12.714	2012-01-27 09:37:12.717	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:12.816	2012-01-27 09:37:12.800	2012-01-27 09:37:12.807	2012-01-27 09:37:12.813	2012-01-27 09:37:12.814	2012-01-27 09:37:12.815	100.837	NA	NA	NA
2012-01-27	EURJPY	09:37:12.913	2012-01-27 09:37:12.900	2012-01-27 09:37:12.906	2012-01-27 09:37:12.911	2012-01-27 09:37:12.911	2012-01-27 09:37:12.912	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:13.014	2012-01-27 09:37:13.000	2012-01-27 09:37:13.007	2012-01-27 09:37:13.011	2012-01-27 09:37:13.012	2012-01-27 09:37:13.013	NA	100.859	NA	NA
2012-01-27	EURJPY	09:37:13.119	2012-01-27 09:37:13.100	2012-01-27 09:37:13.110	2012-01-27 09:37:13.116	2012-01-27 09:37:13.117	2012-01-27 09:37:13.118	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:13.215	2012-01-27 09:37:13.200	2012-01-27 09:37:13.207	2012-01-27 09:37:13.213	2012-01-27 09:37:13.213	2012-01-27 09:37:13.215	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:13.414	2012-01-27 09:37:13.400	2012-01-27 09:37:13.406	2012-01-27 09:37:13.411	2012-01-27 09:37:13.411	2012-01-27 09:37:13.412	NA	100.840	NA	NA
2012-01-27	EURJPY	09:37:13.515	2012-01-27 09:37:13.500	2012-01-27 09:37:13.507	2012-01-27 09:37:13.513	2012-01-27 09:37:13.513	2012-01-27 09:37:13.514	100.840	100.859	100.84	NA
2012-01-27	EURJPY	09:37:13.614	2012-01-27 09:37:13.600	2012-01-27 09:37:13.607	2012-01-27 09:37:13.612	2012-01-27 09:37:13.613	2012-01-27 09:37:13.614	NA	100.858	NA	NA
2012-01-27	EURJPY	09:37:13.714	2012-01-27 09:37:13.700	2012-01-27 09:37:13.706	2012-01-27 09:37:13.711	2012-01-27 09:37:13.711	2012-01-27 09:37:13.714	100.838	NA	NA	NA
2012-01-27	EURJPY	09:37:13.810	2012-01-27 09:37:13.800	2012-01-27 09:37:13.805	2012-01-27 09:37:13.808	2012-01-27 09:37:13.808	2012-01-27 09:37:13.809	100.837	NA	NA	NA
2012-01-27	EURJPY	09:37:13.910	2012-01-27 09:37:13.900	2012-01-27 09:37:13.905	2012-01-27 09:37:13.908	2012-01-27 09:37:13.908	2012-01-27 09:37:13.909	NA	NA	NA	NA
2012-01-27	EURJPY	09:37:14.013	2012-01-27 09:37:14.000	2012-01-27 09:37:14.006	2012-01-27 09:37:14.011	2012-01-27 09:37:14.011	2012-01-27 09:37:14.012	NA	100.857	NA	NA
2012-01-27	EURJPY	09:37:14.114	2012-01-27 09:37:14.100	2012-01-27 09:37:14.106	2012-01-27 09:37:14.112	2012-01-27 09:37:14.112	2012-01-27 09:37:14.113	NA	NA	NA	NA

Figure 3.2: An extract from the FOREX original dataset. Only the most relevant columns for this thesis work are displayed. NA values indicate that the value in that column has not changed since the previous snapshot (for *bid* and *ask*) or that no trades have occurred since the last snapshot (for *paid* and *given*). See Table 3.1 for a description of the different times reported.

- **Quote information** Each snapshot reports the ten best *bid* levels and the ten best *ask* levels at the time when the snapshot was taken.
- **Trades information** The last price traded before the snapshot was taken is also reported. If the market order that originated the transaction hits the ask, the trade is reported as *paid* (the aggressor is buying), while if it hits the bid (the aggressor is selling) it is referenced to as *given*. [3]

It is important to stress that the snapshot at, say, time t_2 reports only the difference in the state of the order book from the snapshot at t_1, as observed at t_2. As a consequence, the information on the path of the state of the order book between two snapshots is lost. For example, we cannot know if between t_1 and t_2 the best ask has changed only once or a hundred times. Also, if between t_1 and t_2 the ask change many times but at t_2 is again equal to the value at t_1, no record is kept.

As we said, the original dataset contains a huge amount of data and a filtering procedure was necessary in order to extract only the part we used in this work. The data processing we made can be divided in two steps:

- Extract relevant rows and columns;

[3] Note that in FX market the notion of buying and selling is always specified with respect to the base currency. For example, considering the EUR/USD cross, where the base currency is USD, *buying* means to buy EUR in exchange for USD, that is selling USD and vice versa.

Label	Meaning	Example
time	Data was written on HSBC database	08:28:35.308
ebsReferenceTime	EBS time ID for the snapshot	08:28:35.300
ebsMarketUpdateTime	EBS took the snapshot	08:28:35.304
feedHandlerReceiveTime	HSBC feed handler received the snapshot	08:28:35.307
feedHandlerPublishTime	HSBC feed handler published the snapshot on his infrastructure	08:28:35.307
eventCaptureTime	HSBC database infrastructure captured the snapshot	08:28:35.308

Table 3.1: Description of the different time stamps found in the original database. ebsReferenceTime always increases of integer multiples of 100 ms.

- Transform and reorganize the reduced dataset.

As reference times we kept *ebsMarketUpdateTime* and *ebsReferenceTime*, the first is the time at which the snapshot is taken, so it is natural to associate a change in the order book to it. The second is useful as a control time as it always increase by an integer amount of 100 ms and it is also used in the randomization procedure that will be described later. We retained also the best bid and best ask columns, while all the others are discarded.

Since snapshots are taken if a change in at least one of the quotes or trades variables happens, once we cut all the irrelevant columns we have many rows in which none of the columns we retained changes. We thus remove all the identical rows at this point. The final product of the first step a dataset with the times at which a change in either the best ask or the the best bid took place. We refer to it as best-quote-changes dataset. Note that best-quote changes are not exactly the same as mid-quote changes. In fact, since the mid-quote is $M = (B + A)/2$, if bid and ask change by opposite quantities the mid-quote remains the same. This concludes the first step.

The second step's main task is to transform the times in a suitable format. Times are given as character strings with millisecond accuracy. They were first converted in POSIX format using R's built-in functions `strptime` and `as.POSIXct`. Market activity follows human habits (working hours, lunch time, etc.) so it is better to use a civil time instead of UTC, which does not have a Daylight Saving Time. The civil time of our choice is London local time. Since FOREX market is active 24 hours a day, the choice of the civil time is in principle irrelevant. However, London is the main center for FOREX trading, hence London local time is the most natural selection. This is also connected to the currency cross that we observe, if we had only AUD/NZD (Australian Dollar on New Zealand Dollar) this could not be the optimal choice.

UTC coincides with London time only for a part of the year. In fact from the last Sunday of March to the last Sunday of October BST[4] is the official London time. So for the days between 25th March 2012 and 28th October 2012, Mondays start at 23:00 UTC of the previous Sunday and so on. We first concatenated all days and then we re-separated them, this time according to London time. This does not create problems since the time change at 01:00 of Sunday were no trades are normally recorded. Note that we do not change the time values (i.e. they are still UTC), but simply rearrange them.

Finally, although FOREX trading is active 24h a day 7 days a week, the activity on Saturdays and Sundays recorded in our dataset is almost negligible. To be more precise, Saturdays are completely empty, while records on Sundays start usually at 18:00 (the time at which the activity in the Far East begins). We therefore eliminated all weekends from the dataset before proceeding further. In the end, for each pair, we have a dataset with best-quote changes events adjusted for Daylight Saving Time.

We used only complete weeks (Monday-Friday) and in total we have 50 weeks starting with Monday 2 January 2012 and ending Friday 14 December 2012, 250 days in total.[5]

3.1.3 News dataset

The activity in all financial markets is sensitive to external news announcements such as those relative to the earnings of listed companies or to the general performance of the economy. The FOREX market is especially influenced by these announcements and in this thesis we present a model that deals with them. We will often refer to these economic data release as "news", though it should be clear that they do not come at unexpected times. Instead their publishing is scheduled well in advance.

We obtained the list of all economic announcements of 2012 from the website *www.dailyfx.com*. The news data consist of a table that reports date and time (UTC) of the news release, the currency most affected by the news, a brief description, an estimate of the importance (high, medium, low), and, where

[4] British Summer Time, BST=UTC+1

[5] Due to an error in the original dataset, the days 6-7-8 August are empty for EUR/JPY. Hence this dataset has actually 247 days.

Date	Time	Time.Zone	Currency	Description	Importance	Actual	Forecast	Previous
Wed Jan 04	09:30	GMT	gbp	GBP Mortgage Approvals	Medium	52.9K	52.8K	52.8K
Wed Jan 04	09:30	GMT	gbp	GBP M4 Money Supply (MoM)	Low	-0.6%		-0.3%
Wed Jan 04	09:30	GMT	gbp	GBP M4 Money Supply (YoY)	Low	-2.6%		-2.7%
Wed Jan 04	10:00	GMT	eur	EUR Euro-Zone Consumer Price Index Estimate (YoY)	Medium	2.8%	2.8%	3.0%
Wed Jan 04	10:00	GMT	eur	EUR Italian Consumer Price Index (NIC incl. tobacco) (MoM)	Low	0.4%	0.3%	-0.1%
Wed Jan 04	10:00	GMT	eur	EUR Italian Consumer Price Index (NIC incl. tobacco) (YoY)	Low	3.3%	3.2%	3.3%
Wed Jan 04	10:00	GMT	eur	EUR Italian Consumer Price Index - EU Harmonized (MoM)	Low	0.3%	0.2%	-0.1%
Wed Jan 04	10:00	GMT	eur	EUR Italian Consumer Price Index - EU Harmonized (YoY)	Low	3.7%	3.5%	3.7%
Wed Jan 04	12:00	GMT	usd	USD MBA Mortgage Applications	Low	-4.1%		0.3%
Wed Jan 04	15:00	GMT	usd	USD Factory Orders	Medium	1.8%	2.0%	-0.2%
Wed Jan 04	22:30	GMT	aud	AUD AiG Performance of Service Index	Medium	49.0		47.7
Wed Jan 04	23:50	GMT	jpy	JPY Monetary Base (YoY)	Low	13.5%		19.5%
Thu Jan 05	00:11	GMT	usd	USD Total Vehicle Sales	Low	13.52M	13.50M	13.59M
Thu Jan 05	00:11	GMT	usd	USD Domestic Vehicle Sales	Low	10.45M	10.42M	10.53M
Thu Jan 05	00:30	GMT	aud	AUD Trade Balance (A$)	Medium	1380M	1650M	1418M
Thu Jan 05	02:30	GMT	cny	CNY China HSBC Services PMI	Low	52.5		52.5
Thu Jan 05	05:00	GMT	JPY	JPY Vehicle Sales (YoY)	LOW			24.1%
Thu Jan 05	07:00	GMT	eur	EUR German Retail Sales (MoM)	Low	-0.9%	0.2%	-0.2%
Thu Jan 05	07:00	GMT	eur	EUR German Retail Sales (YoY)	Medium	0.8%	0.7%	-0.6%
Thu Jan 05	07:45	GMT	eur	EUR French Consumer Confidence Indicator	Low	80	80	80
Thu Jan 05	09:00	GMT	eur	EUR Italian Unemployment Rate s.a.	Low	8.1%	8.1%	8.0%
Thu Jan 05	09:00	GMT	eur	EUR Italian Unemployment Rate (SA)	Low	8.6%	8.5%	8.5%
Thu Jan 05	09:30	GMT	gbp	GBP Purchasing Manager Index Services	Medium	54.0	51.5	52.1

Figure 3.3: Extract from the original news dataset. Each row correspond to a certain data announcement and reports date and time of the announcements, the currency to which is more relevant, a brief description, an importance classification (High, Medium, Low), and where given, the actual data that was published, the analysts forecast, and the previous reading.

Date	Time	Time.Zone	Currency	Description	Importance	Actual	Forecast	Previous
Thu Jan 19	13:30	GMT	usd	USD Consumer Price Index (MoM)	Medium	0.0%	0.1%	0.0%
Thu Jan 19	13:30	GMT	usd	USD Consumer Price Index Ex Food & Energy (MoM)	Medium	0.1%	0.1%	0.2%
Thu Jan 19	13:30	GMT	usd	USD Consumer Price Index (YoY)	High	3.0%	3.0%	3.4%
Thu Jan 19	13:30	GMT	usd	USD Building Permits	Medium	679K	679K	680K
Thu Jan 19	13:30	GMT	usd	USD Building Permits (MoM)	Low	-0.1%	-0.2%	5.6%
Thu Jan 19	13:30	GMT	usd	USD Housing Starts	Medium	657K	680K	685K
Thu Jan 19	13:30	GMT	usd	USD Housing Starts (MoM)	Low	-4.1%	-0.7%	9.1%
Thu Jan 19	13:30	GMT	usd	USD Consumer Price Index Ex Food & Energy (YoY)	High	2.2%	2.2%	2.2%
Thu Jan 19	13:30	GMT	usd	USD Consumer Price Index n.s.a.	Medium	225.672	225.782	226.230

Figure 3.4: An example of many indicators published at the same time. 6 out of 8 concerns inflation in the US, the other two the US house market.

available, expected (Forecast) and actual value. The currencies present in the dataset are USD, EUR, JPY, GBP, CHF, AUD, CAD, NZD and CNY. An extract is reproduced in Figure 3.3.

The "Forecast" value reported in the database is an average of analysts expectations. Big financial data providers, such as Bloomberg or Reuters, conduce surveys among economic analysts asking their forecast of the data that will be released, for example their estimate of the number of new jobs that has been created in a certain month. Then the mean (or median) value of the analysts predictions is taken as the so-called *consensus* estimate and provided to markets participants. The consensus estimate is taken as a reference to differentiate between a good (better than expected) or bad news and therefore plays a very important role. Not all the news in the dataset concerns the publication of an indicator. An example are press conferences of central banks' governors. Clearly since there is no numeric value associated to them, there isn't also any forecast.

Since we treat only the EUR/USD, EUR/JPY and USD/JPY crosses, we retained only the news that are linked to one of these currencies. Note that it is reasonable to do this because these are, by far, the most important currencies in the world. If we were dealing with minor currencies, this would have been much more questionable since an important news coming from the United States is likely to have a broader influence than the sole USA, both because of the prominent economic role of the United States and because USD is the most important reserve currency, so its value influences directly that of many other. This is true also for EUR and JPY, though to a lesser extent.

As for the FOREX dataset, we transformed the time in POSIX format. It is important to underline that more than one economic indicator can be published at the same time. Usually they refer, at least in part, to closely related indicators. An example of this is reproduced in Figure 3.4.

The filtering based on the currency and the time format transformation are the only changes made to the original dataset. Any further filtering is discussed when relevant.

3.2 Descriptive analysis of the data

In this section we give an overview of the principal characteristics of the data in our three samples. The first point of the analysis is the so called seasonality of the data. That is, the variation of market activity, measured in number of events per time interval, that is observed during the day (or the week). We then consider other properties of the data such as the distribution of the times between subsequent events. For the equity database we performed only a brief examination, while a more in-depth analysis was conducted on the FOREX data.

3.2.1 Equity data

In all equity markets activity tends to concentrate at market openings and, to a lesser extent at market closing. Lunch time is characterized by a relatively low level of activity. Our data clearly follow this U-shaped pattern as can be seen from Figure 3.5, where both a random day from our sample and a median day are displayed. We constructed the median day dividing each day of the sample in interval of 30 min and calculating the number of trade events in each interval. Then the median day has in each interval the median values calculated across all days. Figure 3.6 shows the distribution of the number of events in two particular intervals across the sample. The distribution is very broad in some cases while in others some days have uncommon values of activity and this motivated the use of the median.

We note that activity concentrates heavily in the first 30 min and then it falls quite abruptly. This clearly has implications for the modelling that will be discussed in Chapter 4.

Table 3.2 reports some descriptive statistics of the data and in Figure 3.7 the observed distribution of inter-arrival times between trades is plotted. We observe that the single most populated bin is the one that ranges from 0 to 25 ms. In particular the durations shorter than 10 ms represent 16.0% of the total and those equal to the time resolution of the data, i.e. 1 ms, are the 5.19%. Although the distribution is peaked at short durations, it has also a long tail. Durations greater than 1 s represent in fact 45% of the total and those greater than 10 s are still 10% of the total. Duration over 100 s are rare (0.04 %). The complement to one of the empirical cumulative distribution function plotted in Figure 3.7 summarizes these observations.

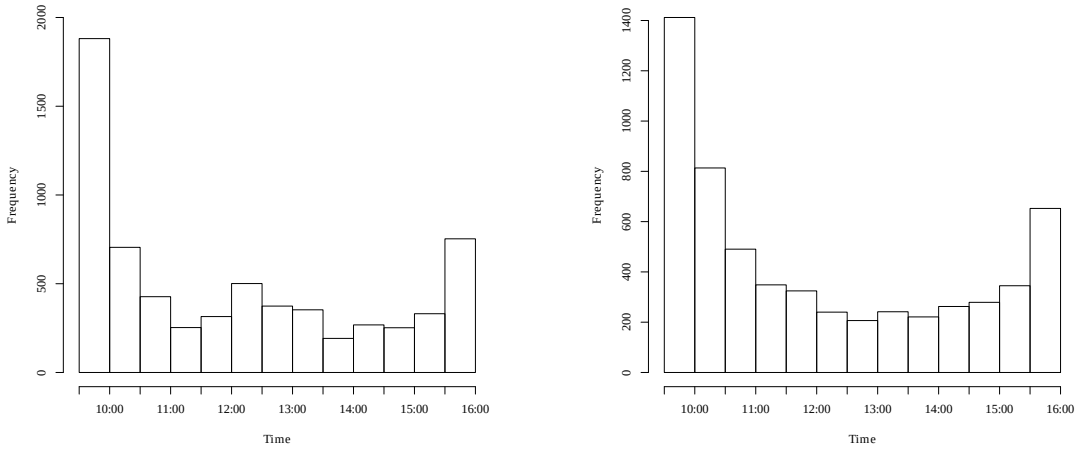

Figure 3.5: Histogram of the number of transactions at different times of day, binned in 30 min intervals. Left: a random day from our sample. Right: median values among all days in the sample.

Number of days	Trades per day (Avg.)	Avg. trade duration (s)	Median trade duration (s)
42	6392	3.66	0.736

Table 3.2: Summary statistics for the Apple stock. Averages and medians are calculated aggregating all the durations.

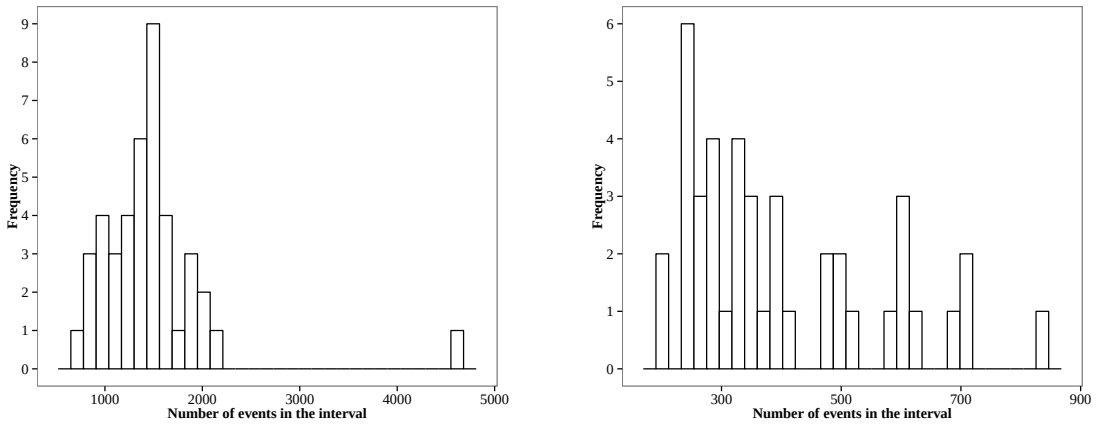

Figure 3.6: Histogram of the number of transactions in the same interval of the day across the different days of the sample. Interval 09:30-10:00 is plotted on the left, interval 11:00-11:30 on the right. There is one day in particular in which a very high activity was observed.

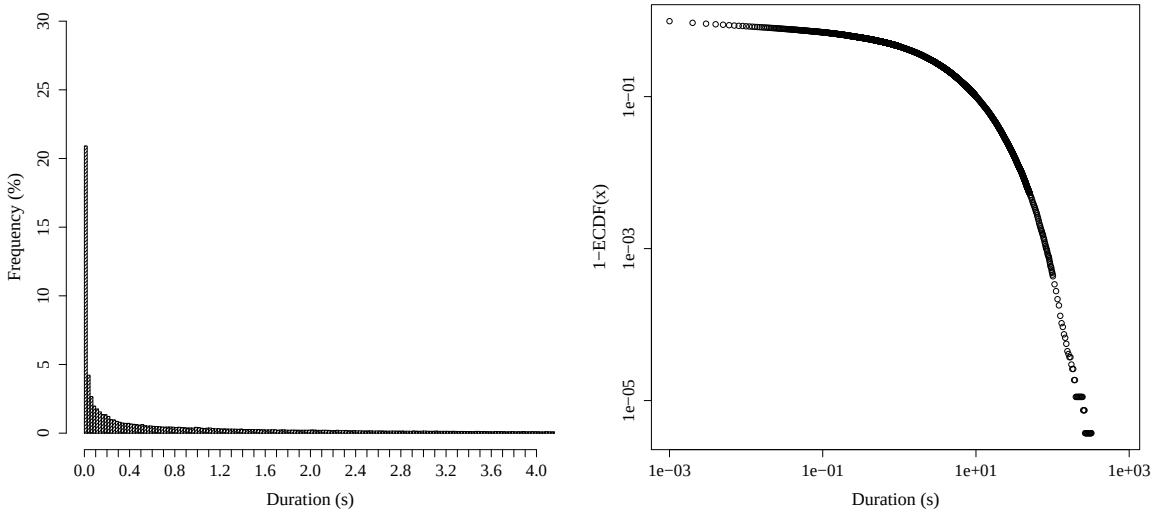

Figure 3.7: Left: histogram of inter-trade duration distribution for the Apple dataset. The bin size is 25 ms, the x-axis is limited for clarity at 4 s, but the maximum observed duration is 323.7 s. Right: Complement of the Empirical cumulative distribution function for the inter-trade duration for the Apple dataset, both scales are logarithmic.

3.2.2 FOREX data

One of the main difference between the FOREX market and other markets is that the former has no specific physical venues and it has no opening or closing times. It is possible to trade 24 hours a day, all days of the week. Hence, the activity pattern is different from that of equity markets. In Figure 3.8 we show the activity on a typical week from our EUR/USD sample. As we explained in Section 3.1.2 we removed Saturdays and Sundays from the database, so here a week means Monday to Friday. Already from Figure 3.8 it emerges that activity varies greatly during the day and, to a lesser extent, also during the week.

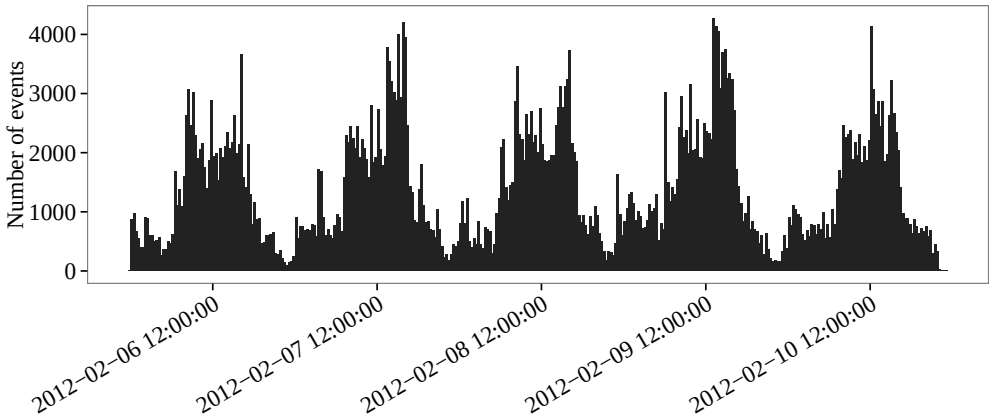

Figure 3.8: Pattern of weekly activity from a typical week of our sample, for EUR/USD data. Counts refer to the number of changes in the best quotes in each interval. The bin size adopted to calculate the number of events is 20 min.

In Figure 3.9 the intraday median pattern for each day of the week and for a given currency pair is displayed. The pattern is estimated through the following procedure:

1. Every day of the sample is divided in intervals of 20 min. The number of quote changes in each interval is recorded.

2. For each day of the week (Mon,Tue, ...), the median of the number of events is calculated among all the corresponding intervals (00:00-00:20, 00:20-00:40, ...)

3. The "median" Monday (Tuesday, ...) is constructed using the median values obtained in step 2.

Figure 3.10 reports instead the plots of the median duration, fitted with a cubic spline across different days of the week. The nodes are the interval lengths divided by the median counts per interval obtained with the procedure outlined above. The pattern of the durations is specular to that of the counts.

Looking at these Figures, we can divide the day in three parts, that correspond to the activity of the major financial centres:

- **Asian Trading** from 22:00 to 07:30 (next day), in this period the activity comes mainly from the Far East financial hubs, particularly Tokyo and Hong Kong.

- **London Trading** from 07:30 to 16:30, we note that this window registers by far the highest level of activity. As we said, London is the main center for FOREX trading both for historical legacy and for geographical reasons.

- **American Trading** from 16:30 to 22:00 activity is sustained by traders based in North America.

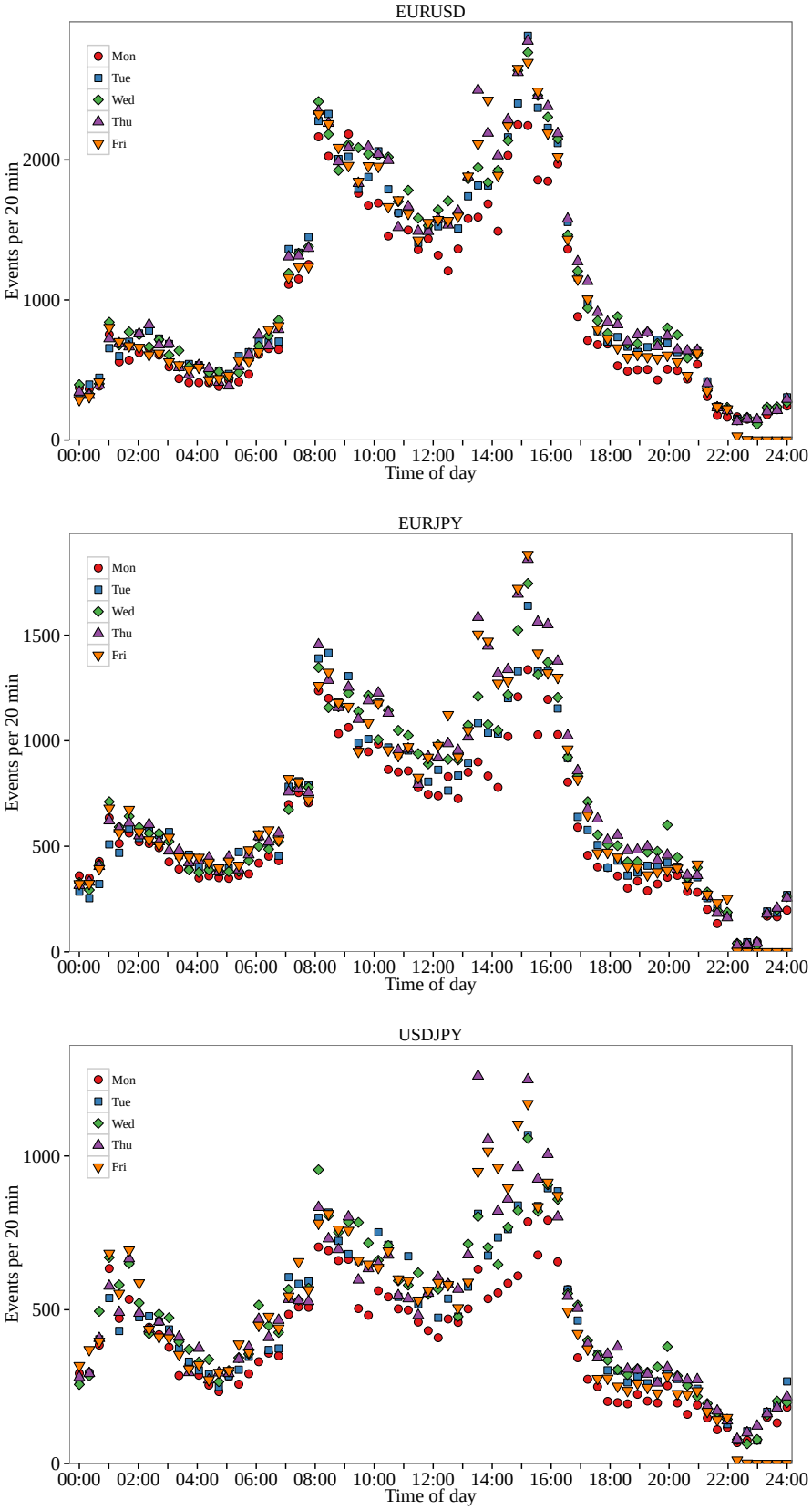

Figure 3.9: Median intraday seasonality for the three different pairs of currencies, for different days of the week. The pattern across different days of the week is almost the same, though the absolute level of activity varies. We can note that the contribution of Asian trading to USD/JPY is higher in relative terms then for the other currencies.

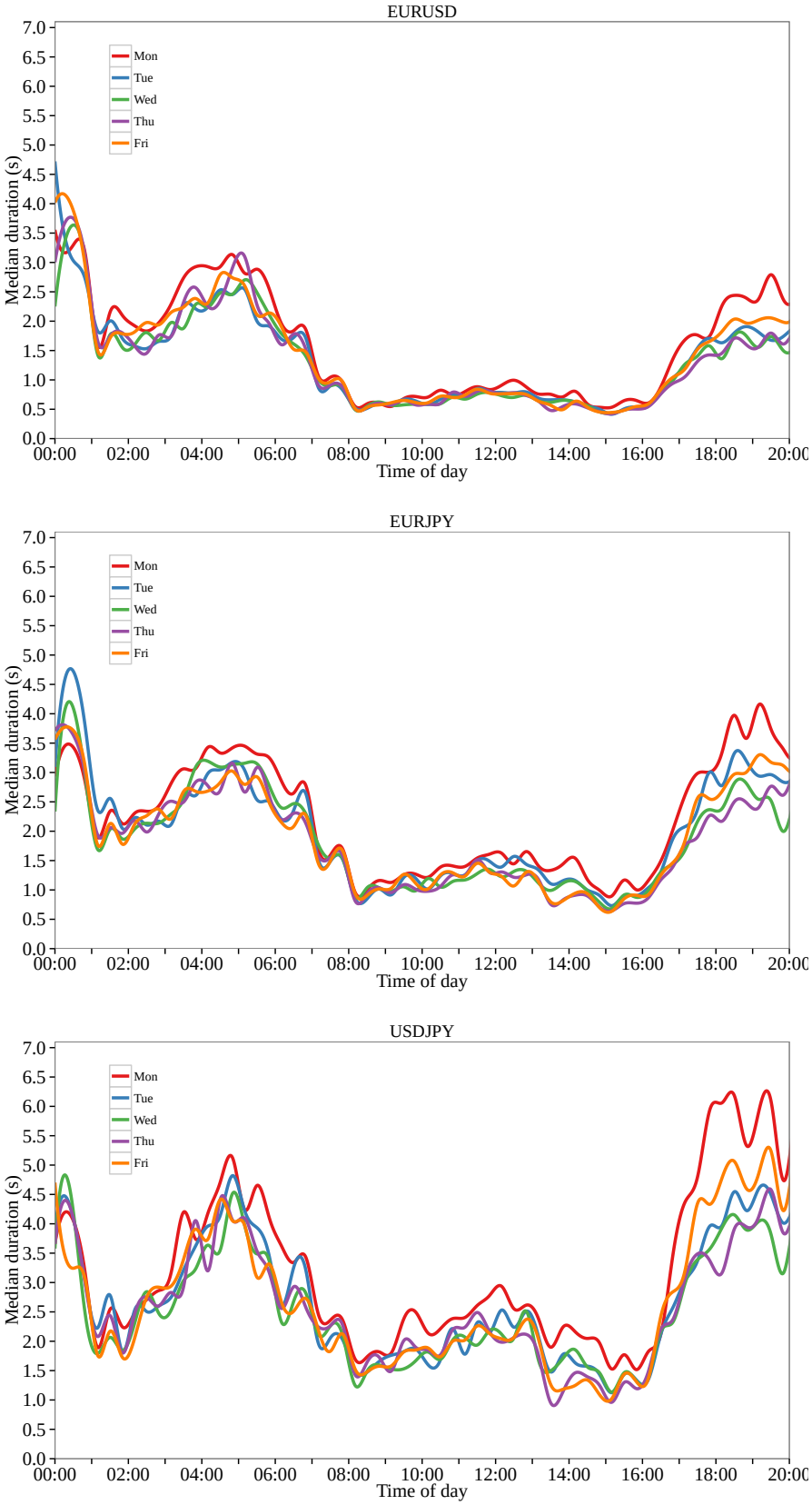

Figure 3.10: Median intraday duration pattern for the three different pairs of currencies, the different level of activity between the three pairs emerges. The functions are not plotted after 20:00 because, as few events happen, the splines fluctuates wildly and give no valuable information.

It is clear that those presented above are not rigid divisions, since trading activity from different areas overlaps at certain times. This is particularly evident in the number of orders submitted during the London afternoon, when there is a peak due to the start of working activity in New York. We observe also that the pattern among different days of the week does not change significantly, apart from the end of Friday when activity stops for the weekend. However, different days of the week are characterized by different levels of activity. This is more evident in Figure 3.11, where the intra-week seasonality is shown. Market activity is more intense between Wednesday and Friday, among other reasons, because most relevant economic announcements, e.g. central banks' interest rate decisions and US unemployment figures, are usually published in these days.

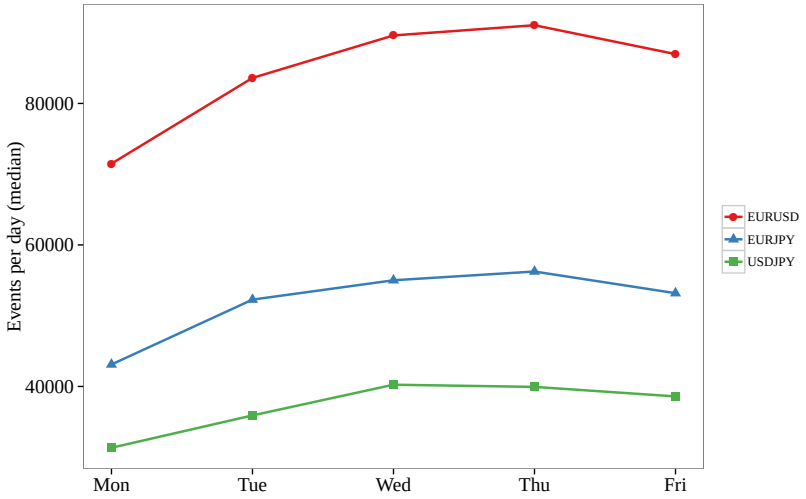

Figure 3.11: Median number of events (mid quote changes) per day of week, when the whole day is considered. Thursdays appear to be the most active days, Mondays the least active ones. We note the different level of activity among the three currency pairs.

To conclude the analysis on seasonality, in Figure 3.12, the median activity on the same day of the week is plotted for the three pairs of currencies. The pattern is almost the same for all the pairs, though the level of activity is quite different, with EUR/USD being the more liquid and USD/JPY the least traded of the three. Also, the relative contributions from different geographical areas have different weights. Although London working hours are the most active in every case, for USD/JPY Asian Trading plays a relevant role.

This examination of the seasonality pattern was important in deciding which data to use in testing our models. It emerged clearly that using the whole day without adjustment can lead to highly flawed results, given the presence of such a clear pattern. For these reasons we decided to restrict our work to London working hours (07:30-16:30). Therefore all the analysis from now on is conduced on this reduced sample window. Table 3.3 reports the average number of events during London working hours for the three currency pairs. Since on 23 September 2012 the tick size on EBS market was increased for all currency pairs, summary statistics are reported separately for the two periods.

With this in mind, we now study the properties of the inter-arrival times between changes in the best quotes. Histograms of the observed durations, calculated from *ebsMarketUpdateTime*, for each currency are plotted in Figure 3.13, the bin size is 25 ms and intervals are right-closed. For all the pairs, about 40 % of the durations are comprised in the interval $(0.075, 0.125]$s. Time intervals longer than 100s are observed but they are rare ($< 1 \times 10^{-3}$ % for EUR/USD, $< 5 \times 10^{-2}$ % for EUR/JPY and USD/JPY).

From the histograms it emerges that the lengths of the time intervals are clustered around integer multiples of 100 ms. This is due to the book snapshots being updated (at most) every 100 ms. We note also that this rule seems not to be strictly enforced, because durations shorter than 100 ms are observed.An histogram of the durations zoomed around 100 ms is shown in Figure 3.14 for EUR/USD, the bin size is 1 ms. The

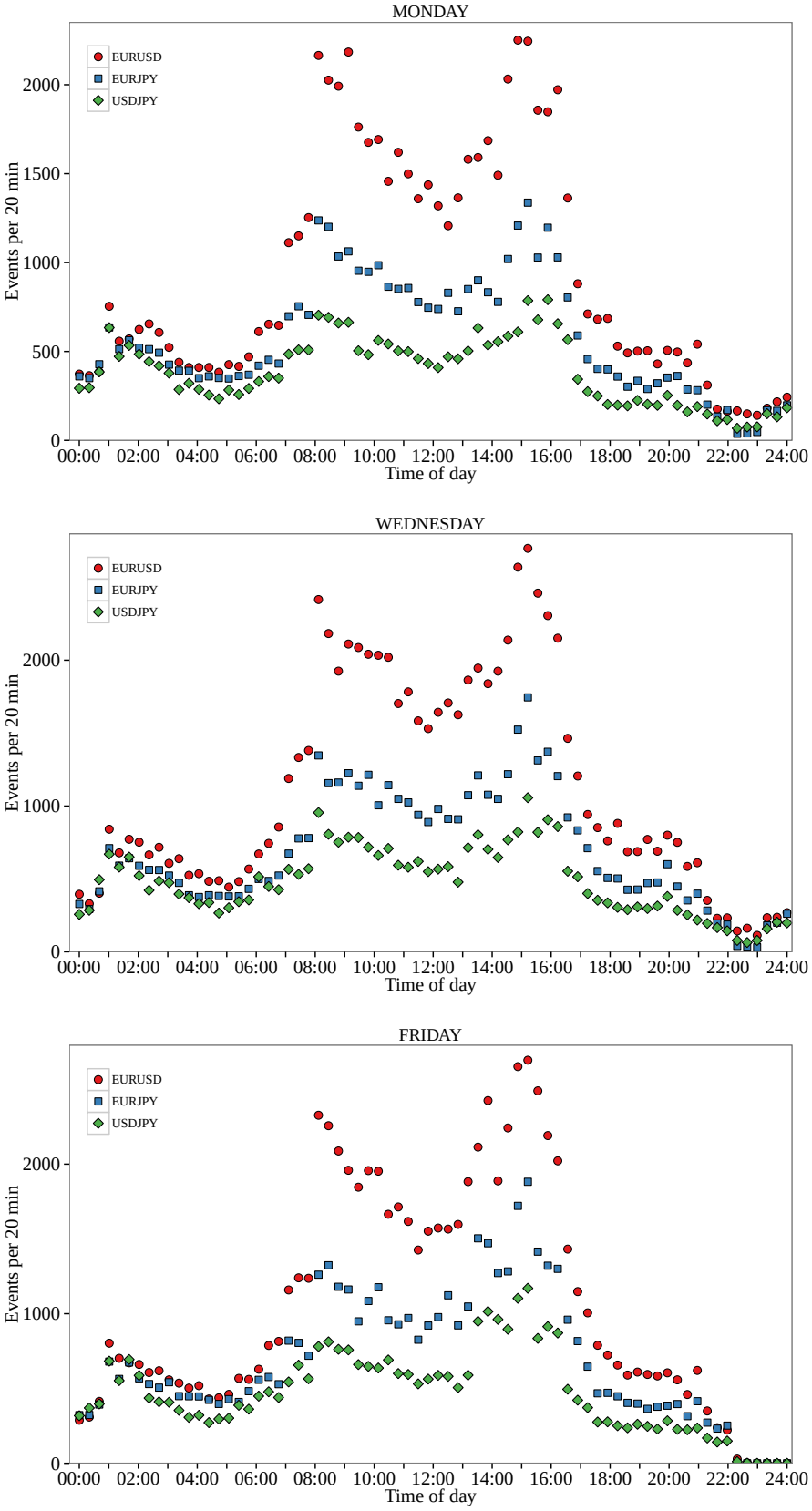

Figure 3.12: Comparison between different pairs on the same day of week. Apart from a clear difference in trading activity, the shape of the pattern is almost the same.

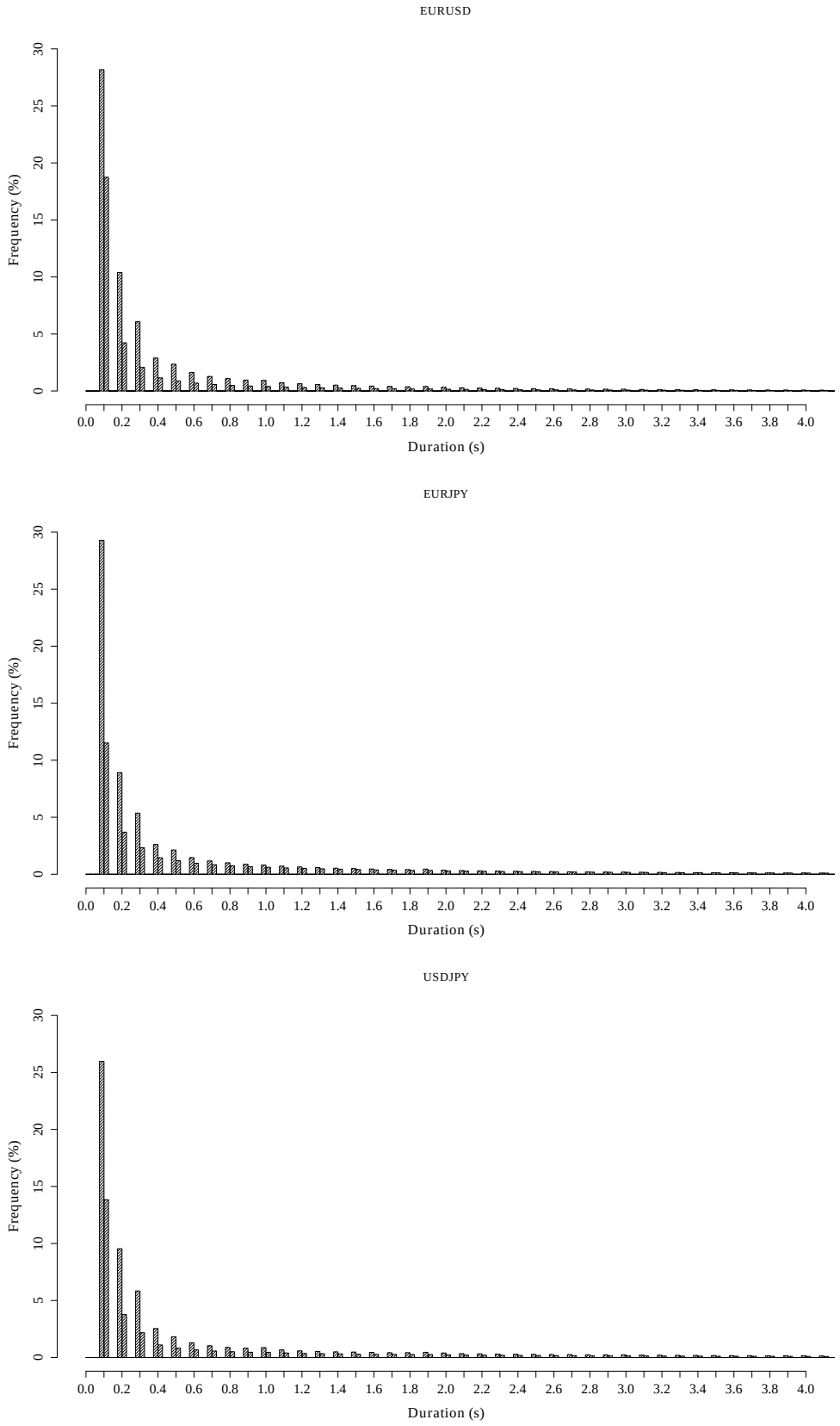

Figure 3.13: Histograms of the durations for each currency pair.

	Number of days	Events per day (Avg.)	Avg. quote duration (s)
		Small tick size	
EUR/USD	190	58902	0.55
EUR/JPY	190	34857	0.91
USD/JPY	190	24477	1.32
		Large tick size	
EUR/USD	60	20229	1.60
EUR/JPY	60	19012	1.70
USD/JPY	60	7244	4.47

Table 3.3: Summary statistics for the FOREX dataset during London working hours, before and after the tick size increase.

number of differences between subsequent ebsMarketUpdateTime stamps lower than 0.05 s is negligible, only 26 for EUR/USD, 10 and 4 for EUR/JPY and USD/JPY respectively. They are probably due to errors in the database. Instead, values between 0.095 and 0.1 s are common. The presence of these time differences lower than 100 ms can be attributed to rounding errors both in the original database and in the conversion we made. In no case, however, the difference between the corresponding *ebsReferenceTimes* is less than 100 ms.

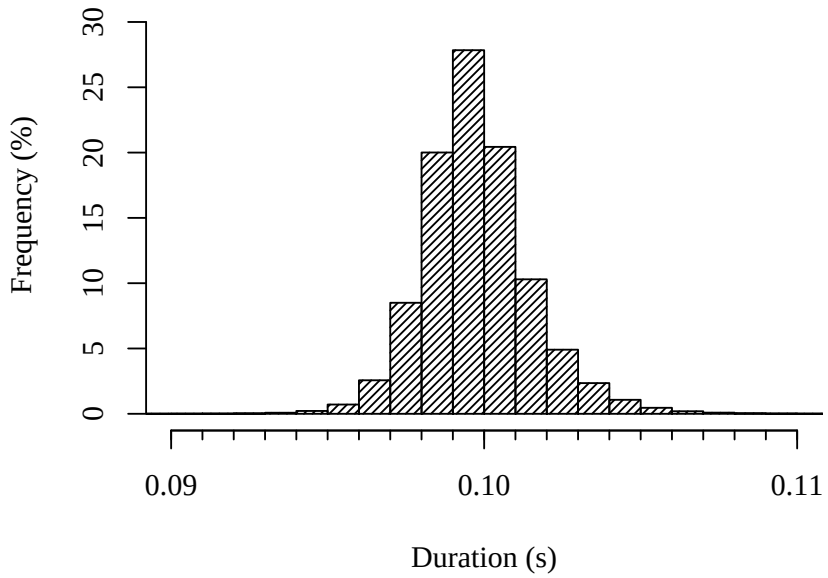

Figure 3.14: Histogram of the duration distribution around 100 ms for EUR/USD. The bin size is 1 ms and the frequency is calculated with respect to the sole durations in the interval $(0, 0.15]$ s and not to the whole sample. The distribution for the other pair is similar.

The presence of these durations slightly lower than the 100 ms declared limit should not represent a problem for our purposes. A more serious issues is the clustering of the durations around multiples of 100 ms, this and possible remedies will be discussed in Chapter 5.

Having discussed the time properties of the data, we now concentrate on those of the quotes.

Although in principle the value of a given currency in terms of another one could be specified with as many digits as one wishes, in practice there exists a minimum value for the increments, the *tick size*. In general, this value depends on the market and on the pair considered. In our case, on EBS platform, the tick size for EUR/USD,EUR/JPY, and USD/JPY were 10^{-5}, 10^{-3} and 10^{-3} until 23 September 2012,

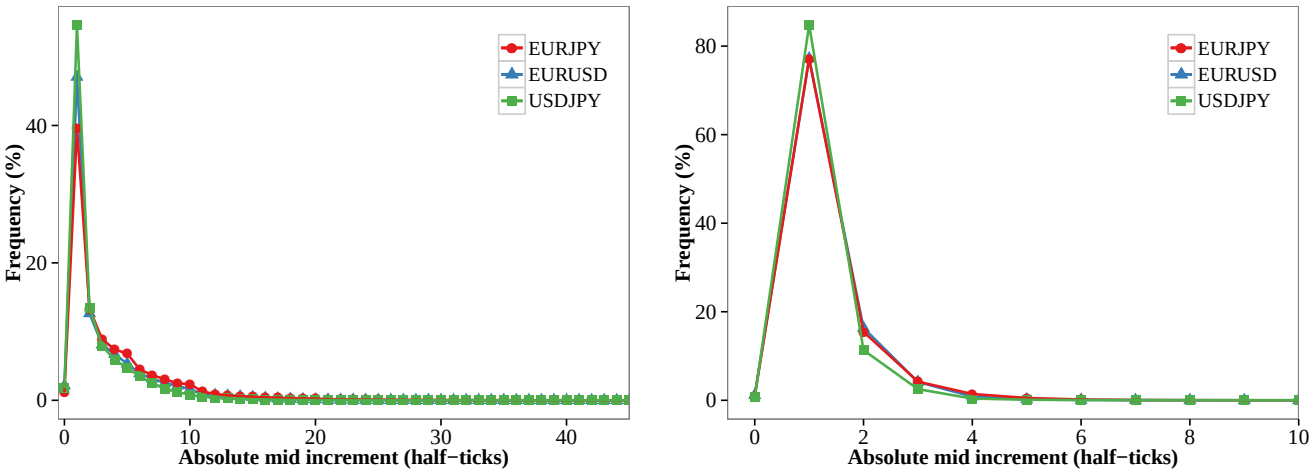

Figure 3.15: Distribution of the mid quote changes (absolute value) measured in mid quote ticks (1 mid tick $= 1/2$ market tick size) for the three currency pairs, before (left) and after (right) the tick size change.

then they were increased to 5×10^{-5}, 5×10^{-3} and 5×10^{-3} [6].

It is important to note that for every possible currency cross there is always a base currency. For example, for the EUR-USD pair the base currency is USD, the tick size is thus expressed in USD in this case, and is not defined for the inverse ratio USD/EUR. For this reason only one ratio per pair is traded and it is the same in all FOREX trading platforms.

It is of some interest then to investigate how the increments (positive and negative) in the mid quote are distributed and what are the typical values of the bid-ask spread. Figure 3.15 shows the absolute value mid quote changes distribution for the three currency pairs, before and after the tick size increase. The behaviour is similar across the pairs: half tick increments are favoured, representing around $50\,\%$ of the movements in the period prior to the tick size increase and over $75\,\%$ with the bigger tick size. Note that a movement of one half tick in the mid, correspond to a change of one tick in the bid while the ask is unchanged or vice-versa. Movements beyond 20 half-ticks in the period prior to the increase or 5 half-ticks after the increase are rare ($< 1\%$). Nevertheless increments as large as 100 half ticks occur ($< 0.01\%$). We note also that there is a certain number, albeit small ($\approx 1\%$), of events for which the increment of the mid price is zero, that is bid and ask move of the same amount in opposite directions.

In Figure 3.16 the spread distribution in ticks is presented. Prior to the tick increase, for EUR/USD and USD/JPY the most common values were 8 and 9 ticks, while for EUR/JPY 20 ticks was the most frequent. A secondary peak is present in EUR/USD distribution at 12 ticks. The large representation of values of the spread close to, but smaller than 10 and 15 in EUR/USD and USD/JPY can be linked to the fact that traders seems to prefer "integer" (or "half integer") values of the price (see Figure 3.17), i.e. ending in 0 or 5. The idea is then that liquidity tends to accumulate at these integer values, and traders anticipate this accumulation and take priority offering slightly better prices.

With the advent of the larger tick size, EUR/USD and USD/JPY spread concentrated at 1-2 ticks, while EUR/JPY rescaled to 4 ticks. EUR/JPY presents in both cases a wider spread, this may be partly due to the fact that the tick size for this pair is the same of USD/JPY, but the values of EUR/JPY rate varied in the range $[94; 111]$ while for USD/JPY the range was $[76; 84]$, thus a tick represents a smaller percentage of EUR/JPY value. However this cannot be the only explanation since the ratio between spread in base currency and price for EUR/JPY is twice as big as that of USD/JPY (0.02% vs. 0.01%).

From Figure 3.16 we see that there is a small fraction of spread whose value is negative ($< 0.1\%$). This in principle is very strange since with a negative spread there is an arbitrage opportunity: buying at the ask

[6] Note that as for a unit change in, say, the ask the mid quote moves by one half, the tick size for the mid quotes is half these values.

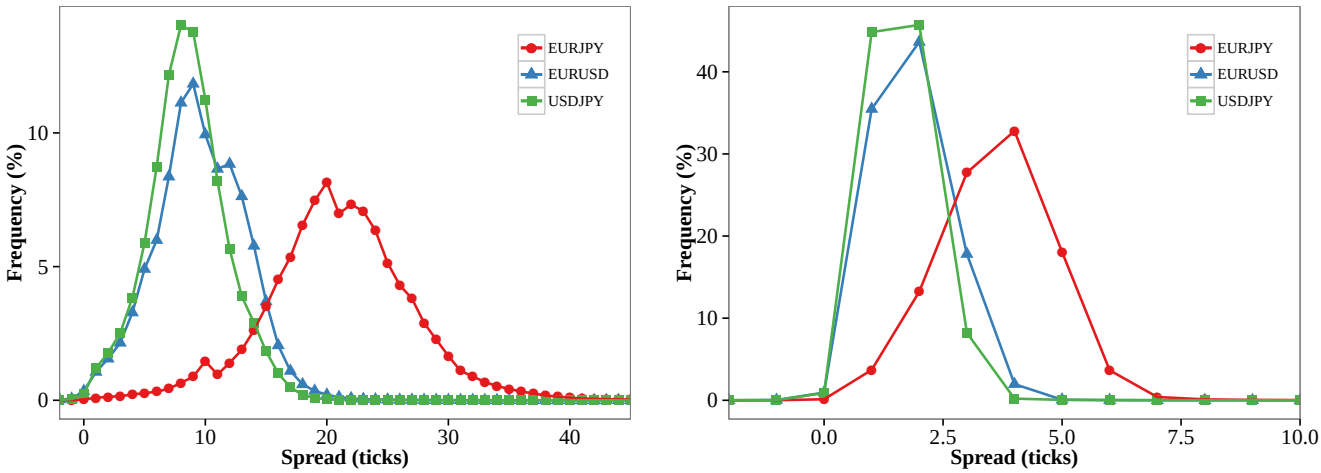

Figure 3.16: Distribution of the spread measured in ticks for the three pairs, before(left) and after (right) the tick size increment. EUR/USD and USD/JPY have similar patterns, while that of EUR/JPY is clearly different.

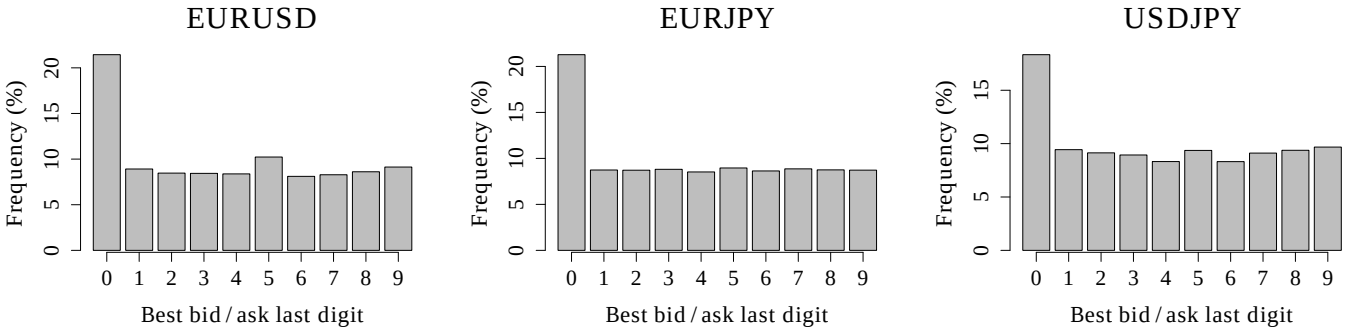

Figure 3.17: Last digit distribution for the three pairs before the tick size increase.

and selling immediately at the bid one makes a risk-less profit. However, in most cases this is not actually possible and it is due to the fact that the quotes on EBS platform are not *credit-checked*. This in essence means that not all the traders can actually trade on a certain quote, they can only if they have a credit line with (i.e. they are connected with) the trader that posted the quote. In any case the discrepancies that give rise to negative spreads fade very quickly, so this anomalous states do not persist for more than one snapshot time.

3.2.3 News data

We provide here only a brief summary of the news data, discussion on news impact is postponed to Chapter 6.

In the news dataset we have 4437 entry for the period spanned by our FOREX dataset (02 Jan - 14 Dec). The totals divided per currency and importance are reported in Table 3.4.

For our purposes it is more relevant the number of distinct news times, and this is equal to 2039.

	HIGH	MEDIUM	LOW	TOTAL
EUR	258	703	820	1781
USD	184	490	1067	1741
JPY	48	325	542	915
TOTAL	490	1518	2429	4437

Table 3.4: Number of news in each group.

3.3 Software

A large part of this thesis work involved the writing of computer code.The code that has been produced can be divided according to his function into five main classes:

- Data filtering and processing, particularly for the FOREX data.
- Parameter estimation through likelihood maximization.
- Simulation of Hawkes processes.
- Goodness of fit assessment.
- Production of final results and plots.

The code has been written mainly in the *R* programming language (R Core Team, 2012). For the most computationally intensive task, namely the likelihood optimization, the likelihood functions were instead coded in Fortran.

R is a free, high-level programming language, his capabilities can be easily extended through the many freely available packages. It provides many functions for data import and statistical analysis. Moreover, C, C++ and Fortran code can be easily linked at run time to speed up the more demanding calculations. Finally, we took advantage of the parallel computing possibilities offered by the built-in *parallel* package and the *foreach* (Revolution Analytics, 2012) package.

In addition to the already cited packages, for the likelihood maximization we made use of the routines accessible through the *optimx* package (in particular the *nlminb* algorithm) and of the Differental Evolution algorithm contained in the package *DEoptim* (Nash and Varadhan, 2011; Ardia et al., 2012)

R comes also with advanced plotting capabilities, all the figures in this thesis were produced in R, either with the built-in *graphics* package or with the *ggplot2* library (Wickham, 2009).

Chapter 4

Modelling equity data with Hawkes processes

In this Chapter we present the application of Hawkes models to equity data. We discuss the performance of a univariate model to describe the trades arrival process. All the work in this chapter is original, though the model specifications are not dissimilar to those found in the relevant literature.

The aim of the work described in this chapter can be summarized by the following two points:

- Assessment of the performance of a univariate linear Hawkes process in describing financial activity, and comparison of the results of different choices for the kernel.
- Comparison of our results with those of similar models already discussed in the literature.

We first discuss how to deal with intraday seasonality, then we give the specification and results of the univariate model with three different functional forms for the kernel, namely single exponential, sum of two exponential and power law. Finally we compare the performances of the three specifications.

4.1 Data

The empirical data for this part are those described in Section 3.1.1. We start with the list of recorded transactions times for each day, rescaled so that $t = 0$ coincides with market opening. These data are further transformed to compensate intraday seasonality as described in the following Section.

4.2 Intraday seasonality

We have seen in Chapter 3 that financial data possess a strong intraday pattern. Hence, in order to apply the model to the data, the issue of intraday seasonality has to be addressed first. Two are the more common approaches found in the literature, one is to restrict the analysis to a subset of data for which is reasonable to neglect seasonality effects, i.e. for which the number of transactions per unit time is approximately constant. The other one is to normalize the data by a seasonality function that accounts for time of day effects.

We chose to test the model with both approaches. The first sample is constituted of transactions occurred between 11:00 and 15:00 of each trading day. In this interval the number of transactions per unit time is constant to a good approximation.

The second sample is built starting from the whole trading day bur removing the first and last five minutes, and normalizing the original durations with a cubic spline function which reproduces the intraday pattern of typical duration. We removed the first and the last 5 minutes because opening and closing are peculiar

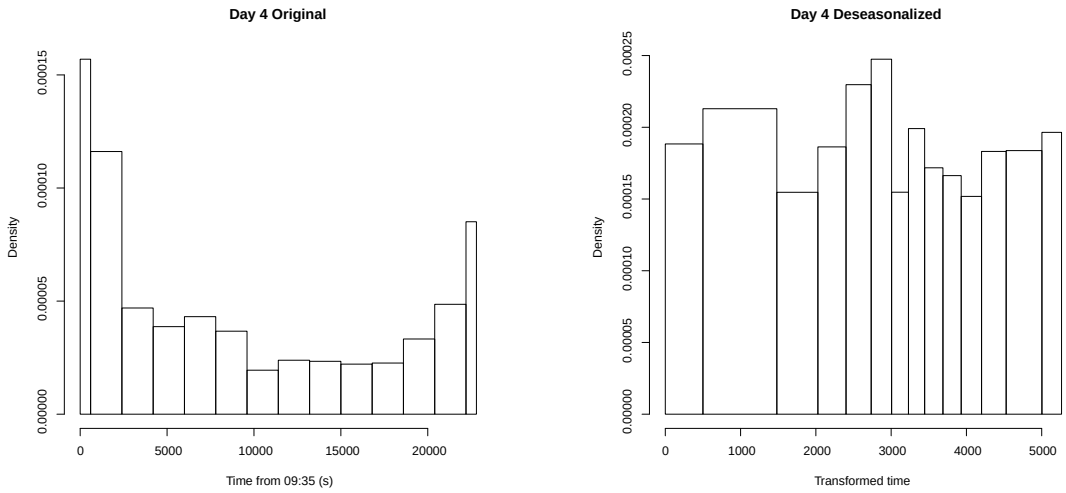

Figure 4.1: Effects of the deseasonalization procedure on a sample day. Left: number of transactions in the intervals described in the procedure on a sample day. Right: same day after the deseasonalization procedure, the classes for the histogram are the transformed intervals.

in that they are transition periods from the initial auction to normal trading and vice versa so they can have different characteristics and alter the analysis of model performance. The procedure we followed is outlined below in more detail:

1. Each trading day is divided in n intervals of length l_i minutes, $i = 1, ..., n$.
2. The number of transactions N_i^D occurred in each interval i and for every trading day D is computed. Then, for each interval the median value among all days, $\overline{N}_i$, is calculated and associated to the center of that interval. In this way a "median" trading day is constructed. For each interval i of this median day the corresponding median duration is computed as: $\overline{d}_i = \frac{l_i}{\overline{N}_i}$.
3. These durations $\overline{d}_i$ are associated to the centres I_i of the corresponding interval.
4. A cubic spline function $s(t)$ is fitted on the points $(I_i, \overline{d}_i)$.
5. Then, for each day, the original durations $d_k = t_{k+1} - t_k$ are transformed according to the formula: $d'_k = \frac{d_k}{s(t_k)}$.
6. Transformed transaction times are eventually obtained from transformed durations as: $t'_{k+1} = t'_k + d'_k$, with $t'_1 = 0$.

The first interval and the last one are chosen to be 10 min long, whereas the remaining intervals last 30 min, so we have a total of 14 intervals. This choice was made as using too many intervals leads to a fluctuating seasonality function, while too few intervals conduce to a poor description of the daily pattern. The shorter intervals placed at the start and at the end of the trading period are motivated by the steep variation in the number of transaction that occur at these times.

The use of the median instead of the mean was prompted by the broad distribution of the number of transaction in each interval across all days. The median, being less sensitive to particular days with very low or very high number of transactions, gives an overall better result.

In the end, we have two samples:

- **11:00-15:00**: 42 four-hour long windows of trading, one for each day. The average number of events in each interval is 2422.

- **Deseasonalized**: transformed trading days. Here we have 5819 events on average. Note that now the times are no more expressed in seconds, but are pure numbers. This will apply also to the estimated parameters.

4.3 General procedure

As discussed in Chapter 2, the maximum likelihood method is the standard procedure for the estimation of the parameters of the Hawkes model. For each type of kernel, we wrote a code for the evaluation of $\ln \mathcal{L}$. A standard maximization algorithm was then applied to find the parameters corresponding to the maximum. The procedure is applied separately for each day, so that the final result consists of a list with the best estimates for each day.

We tested different maximization algorithms available in R, and finally opted for the use of two quasi-Newton routines, namely function *optim* that use the BFGS method (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) and function *nlminb*, which rely upon PORT routines (Gay, 1990).

The estimation procedure that we implemented is as follows: we start with a list each element of which is made of transactions data from a certain day - either those relative to the subinterval 11:00-15:00, or the deseasonalized ones. Once a specific kernel is chosen, the optimization routine is applied to the appropriate likelihood function for each element of the aforementioned list. A set of 10 different starting values for the parameters is used in order to improve the chances of finding the global optimum, and only the result with the higher likelihood is retained. Both maximization routines are tested, and again the results with the higher likelihood are chosen. The uncertainties associated with the best parameters estimates are calculated using the inverse of the Hessian matrix of the log-likelihood evaluated at the maximum, as discussed in Subsection 2.3.2.

4.4 Model details and Results

In the following we specify the model details and we discuss the overall performance of each type of kernel on the two datasets, looking at the results obtained with the procedure outlined above.

4.4.1 Single exponential kernel

We first test a Hawkes process with a simple, single exponential kernel of the form:

$$w_{\text{SE}}(t) = \alpha e^{-\beta t}. \tag{4.1}$$

The intensity function thus reads:

$$\lambda(t) = \mu + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}, \tag{4.2}$$

where μ, α, and β are the constant parameters of the model. μ is a baseline intensity, α sets the amplitude of the self exciting effect, while β fixes the time scale of the influence of a past event on the present intensity. In the single exponential model, we have only one time scale, given by β^{-1}.

The log-likelihood is, from (2.41) and for the interval $[0, T]$:

$$\ln \mathcal{L} = -\mu T - \sum_{t_i} \frac{\alpha}{\beta} \left(1 - e^{-\beta(T-t_i)}\right) + \sum_{t_i} \ln\left(\mu + \alpha R(i)\right), \tag{4.3}$$

$R(i)$ has been defined in (2.40), and we recall that $R(1) = 0$.

In Figure 4.2 and 4.4 we report the distribution across the days of the parameter best estimates for the 11:00-15:00 data and for the deseasonalized data, respectively. For all the parameters the distribution is

11:00-15:00 data			
	μ	α	β
Avg. values (s^{-1})	0.120	9.2	34
Std. dev. ($\%$)	29	38	39
Avg. uncertainity per day ($\%$)	3	9	9
Deseasonalized data			
	μ	α	β
Avg. values	0.79	39	139
Std. dev. ($\%$)	27	30	28
Avg. uncertainity per day ($\%$)	1.7	5	6

Table 4.1: *Single exponential kernel*. Average results across days with corresponding standard deviation. The average uncertainty on the estimates on each day calculated from the Hessian matrix is also given.

quite broad, standard deviations and average values are summarized in Table 4.1. The average uncertainty per day reported in the table is the mean of the uncertainties obtained for each parameter from the inverse of the Hessian matrix of the log-likelihood evaluated at the maximum. These values are much lower than the standard deviation across days, signalling that with the single exponential kernel optimal parameters values are in fact quite different across the days.

The value of β we found in the 11:00-15:00 dataset corresponds to a characteristic time of about 30 ms. In the case of the deseasonalized sample, parameters are pure numbers, and hence the characteristic time has not a direct correspondence in real time. However, we can reason as follows to make a comparison. The transformed day is about 5244 long, this value is obtained as

$$\int_0^{22800} \frac{\mathrm{d}t}{s(t)} \tag{4.4}$$

where $s(t)$ is the spline function described is Section 4.2 and $22\,800\,\mathrm{s}$ is the length of trading day in seconds. This is an approximation since the intervals between events are not really infinitesimal, though they are on average very small. Then, on average, one unit of transformed time corresponds to $22800/5244 = 4.35$ s. So the characteristic time from the deseasonalized sample, $\beta^{-1} \approx 7.2 \cdot 10^{-3}$, is approximately equivalent to $4.35 \cdot 7.2 \cdot 10^{-3} \approx 31$ ms, in agreement to the one found on 11:00-15:00 sample. Transforming in the same way the value of μ and α we obtain $0.18\ \mathrm{s}^{-1}$ and $9.0\ \mathrm{s}^{-1}$ respectively, again in accordance with the results of the 11:00-15:00 dataset.

Figure 4.3 shows quantile-quantile plots of the residuals Λ_i against the standard exponential distribution from two representative days of our sample, chosen among those for which the QQ-plot showed a better and, respectively worse, fit. If the data were exactly described by the model, the residuals would follow a standard exponential distribution. Analogous plots are provided for the deseasonalized data in Figures 4.5. Clearly the fit is not satisfactory. We will comment further the results in the next section.

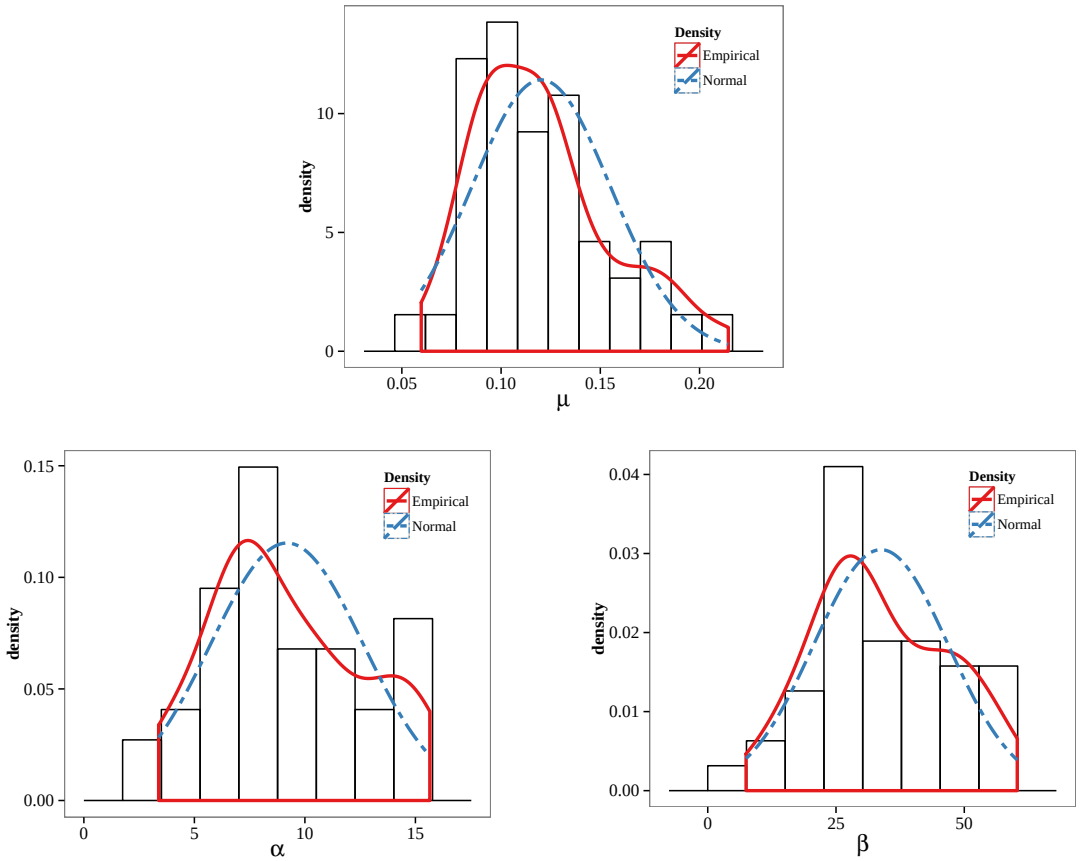

Figure 4.2: *Single Exponential kernel, 11:00-15:00 data.* Distribution of the parameters estimates across the days of the sample. The histogram and the empirical kernel estimates are shown together with the fitted normal density for reference.

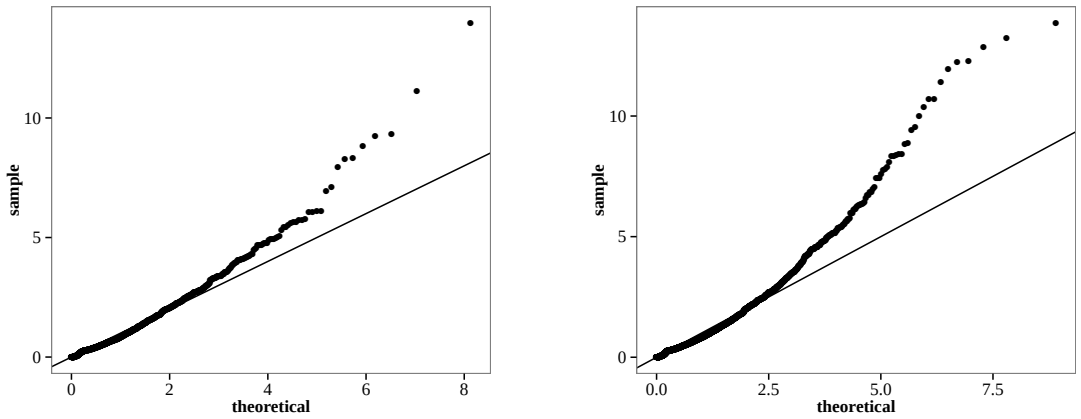

Figure 4.3: *Single Exponential kernel, 11:00-15:00 data.* QQ-plot of residuals against standard exponential distribution for two days obtained with the SE kernel on the 11:00-15:00 transactions data. The graph on the left shows one among the best results, that on the right one among the worst.

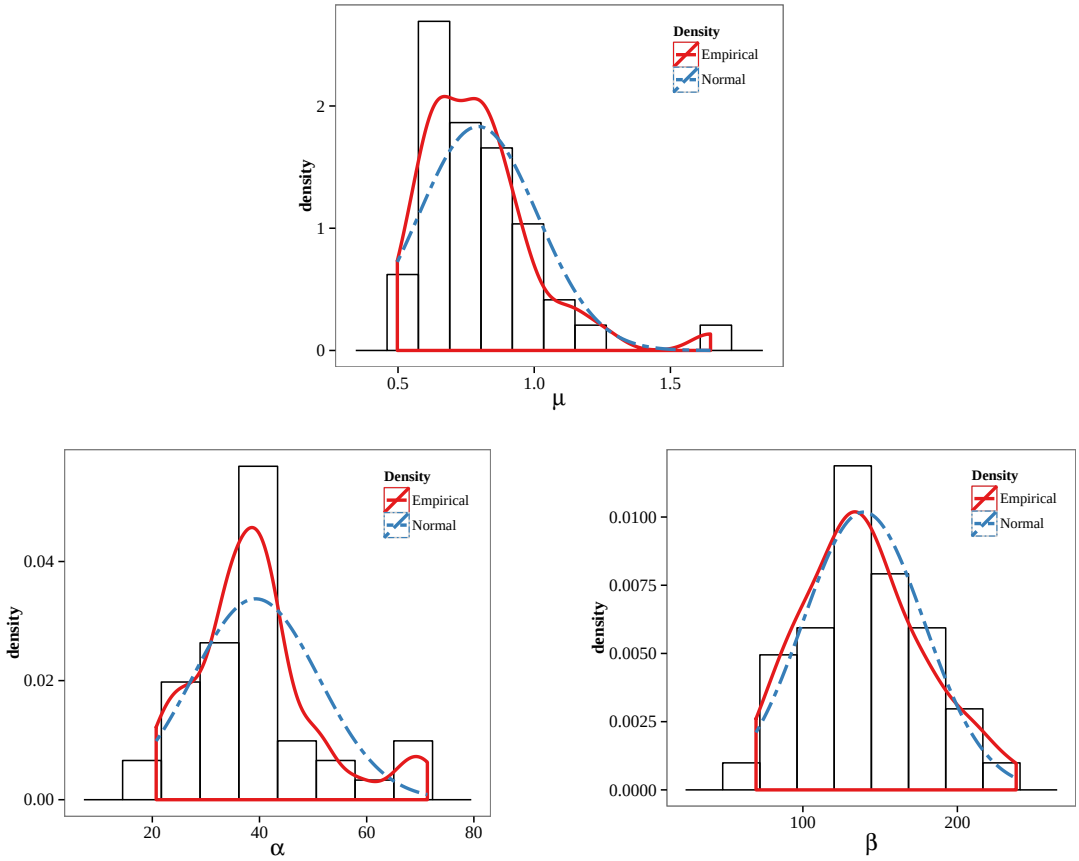

Figure 4.4: *Single Exponential kernel, deseasonalized data.* Distribution of the parameters estimates across the days of the sample. The histogram and the empirical kernel estimates are shown together with the fitted normal density for reference.

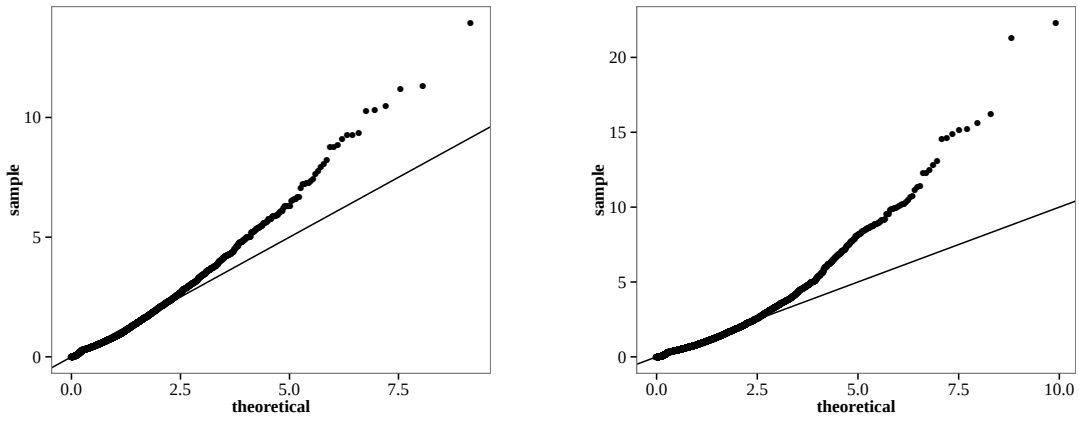

Figure 4.5: *Single Exponential kernel, deseasonalized data.* QQ-plot of residuals against standard exponential distribution for two days obtained with the single exponential kernel on deseasonalized data. The graph on the left shows one among the best results, that on the right one among the worst.

4.4.2 Double exponential kernel

Our second test is with a Hawkes process with a double exponential kernel of the form:

$$w_{\text{DE}}(t) = \alpha_A e^{-\beta_A t} + \alpha_B e^{-\beta_B t}. \tag{4.5}$$

The intensity function is thus written as:

11:00-15:00 data

	μ	α_A	β_A	α_B	β_B
Avg. values (s^{-1})	0.08	19.9	103	0.33	1.6
Std. dev. (%)	32	37	45	84	99
Avg. uncertainity per day (%)	5	10	11	19	21

Deseasonalized data

	μ	α_A	β_A	α_B	β_B
Avg. values	0.58	80	402	1.5	6.0
Std. dev. (%)	25	33	34	52	64
Avg. uncertainity per day (%)	3	6	7	14	16

Table 4.2: *Double exponential kernel.* Average results across days and relative standard deviation. The average uncertainty on the estimates on each day calculated from the Hessian matrix is also reported.

$$\lambda(t) = \mu + \sum_{q=A,B} \sum_{t_i<t} \alpha_p e^{-\beta_p (t-t_i)}, \tag{4.6}$$

where μ, α_q, and β_q ($q = A, B$) are the constant parameters of the model, two more than the single exponential one. Here then two different time scales are present and their relative weight is controlled by the amplitudes α. For definiteness β_A will always be greater than or equal to β_B, i.e. the A subscript refers to the shorter time scale and B to the longer one.

The log-likelihood is, from (2.41) and for the interval $[0, T]$:

$$\ln \mathcal{L} = -\mu T - \sum_{q=A,B} \sum_{t_i} \frac{\alpha_q}{\beta_q} \left(1 - e^{-\beta_q (T-t_i)}\right) + \sum_{t_i} \ln \left(\mu + \alpha_p R_q(i)\right), \tag{4.7}$$

As for the single exponential case, we give here the results of the fit across the days in Figure 4.6 and 4.9. We also provide a summary with the average results in Table 4.2. Again the variability across the sample is pretty high. Especially the parameters α_B and β_B present very large errors and extreme dispersion across the sample.

From the results we can observe that now two distinct timescales emerge. For the 11:00-15:00 sample the shorter time scale β_A^{-1} is of the order of 10 ms and the longer one β_B^{-1} ranges from about 500 ms to over 5 s. We note that the single time scale found with the single exponential stays in between. The two scales have very different weights, with the shorter one being the more relevant.

Similar considerations apply also to the deseasonalized sample. Proceeding as we did for the single exponential kernel, using the coefficient (4.4), we obtain similar characteristic times to the 11:00-15:00 sample, namely $\beta_A^{-1} \cdot 4.35 \approx 10$ms and $\beta_B^{-1} \cdot 4.35$ varies from about 300 ms to about 9 s. Thanks to the increased sample size, the estimates have in this case a lower variability, albeit still hight in absolute terms.

Figures 4.7 and 4.8 show the distribution of the residuals against the exponential distribution for one of the best and one of the worst days in the 11:00-15:00 and deseasonalized samples respectively. We see that in some cases the fit appears to be very good while others present a more pronounced difference between the residuals distribution and the exponential one.

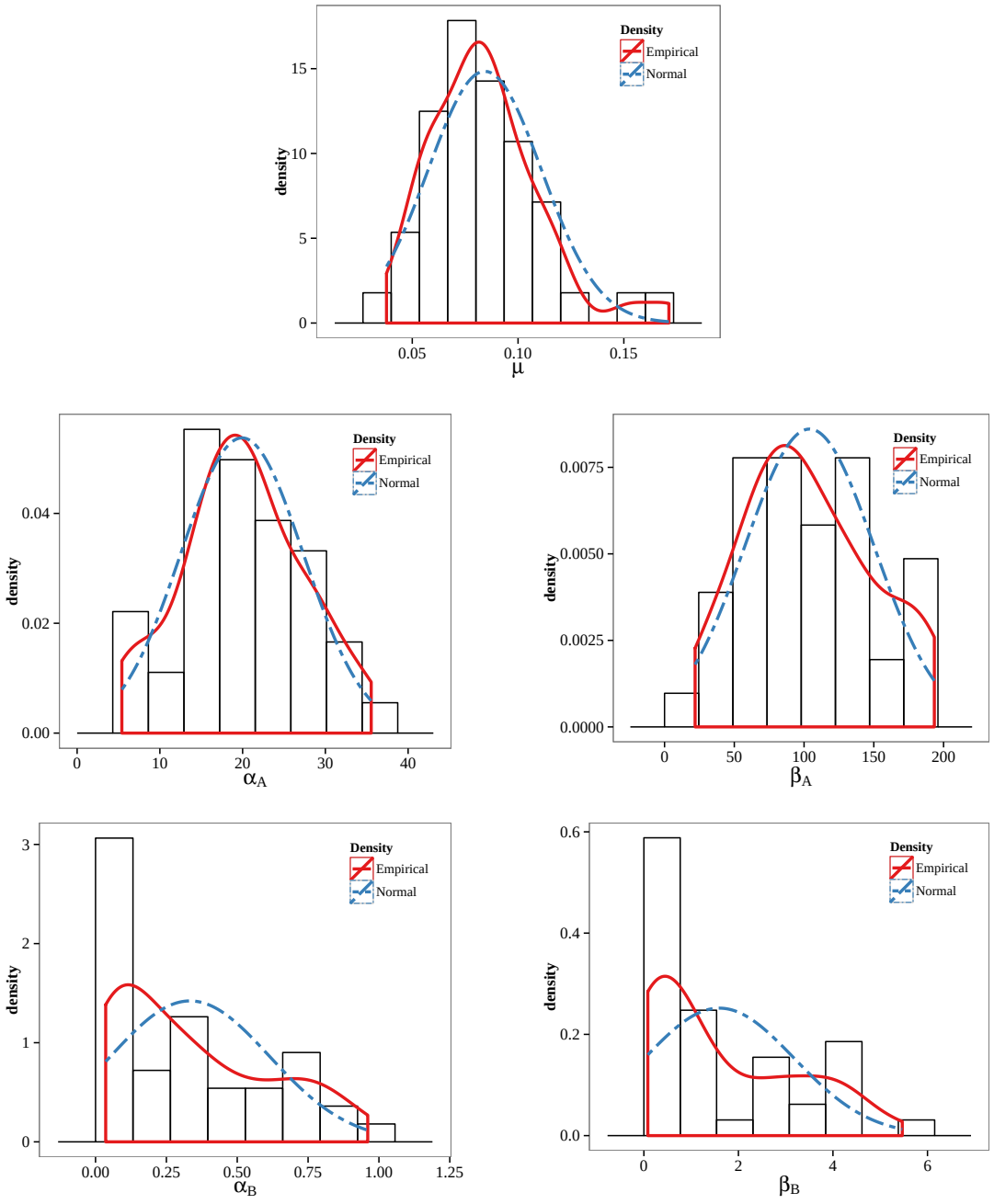

Figure 4.6: *Double Exponential kernel, 11:00-15:00 data*. Distribution of the parameters estimates across the days of the sample. The histogram and the empirical kernel estimates are shown together with the fitted normal density for reference.

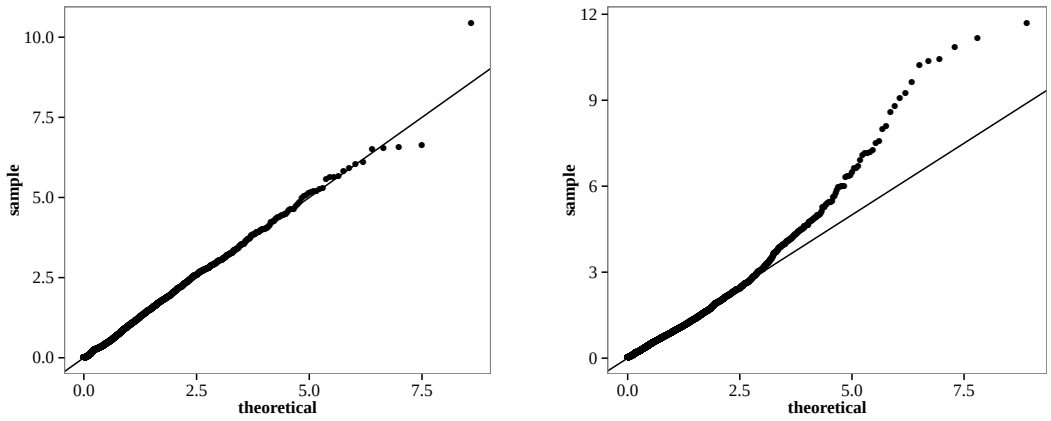

Figure 4.7: *Double Exponential kernel, 11:00-15:00 data.* QQ-plot of residuals against standard exponential distribution for two days obtained with the single exponential kernel. The graph on the left shows one among the best results, that on the right one among the worst.

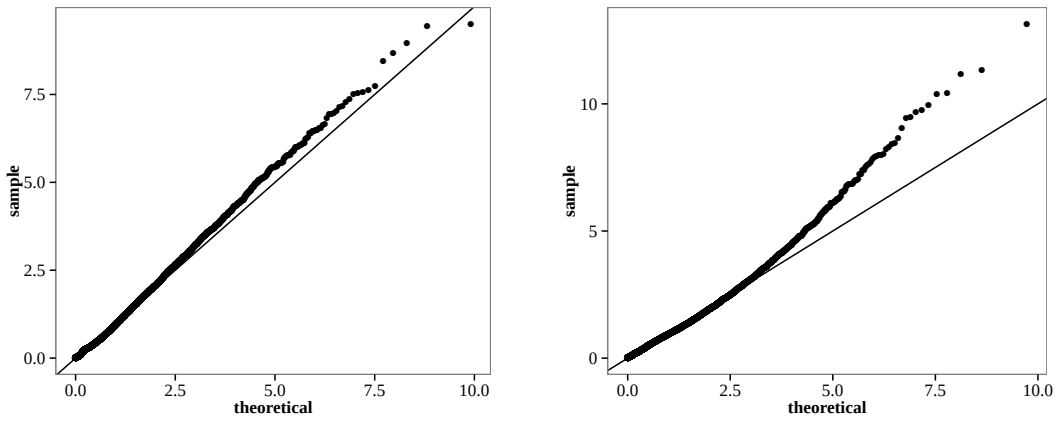

Figure 4.8: *Double Exponential kernel, deseasonalized data.* QQ-plot of residuals against standard exponential distribution for two days obtained with the single exponential kernel on deseasonalized data. The graph on the left shows one among the best results, that on the right one among the worst.

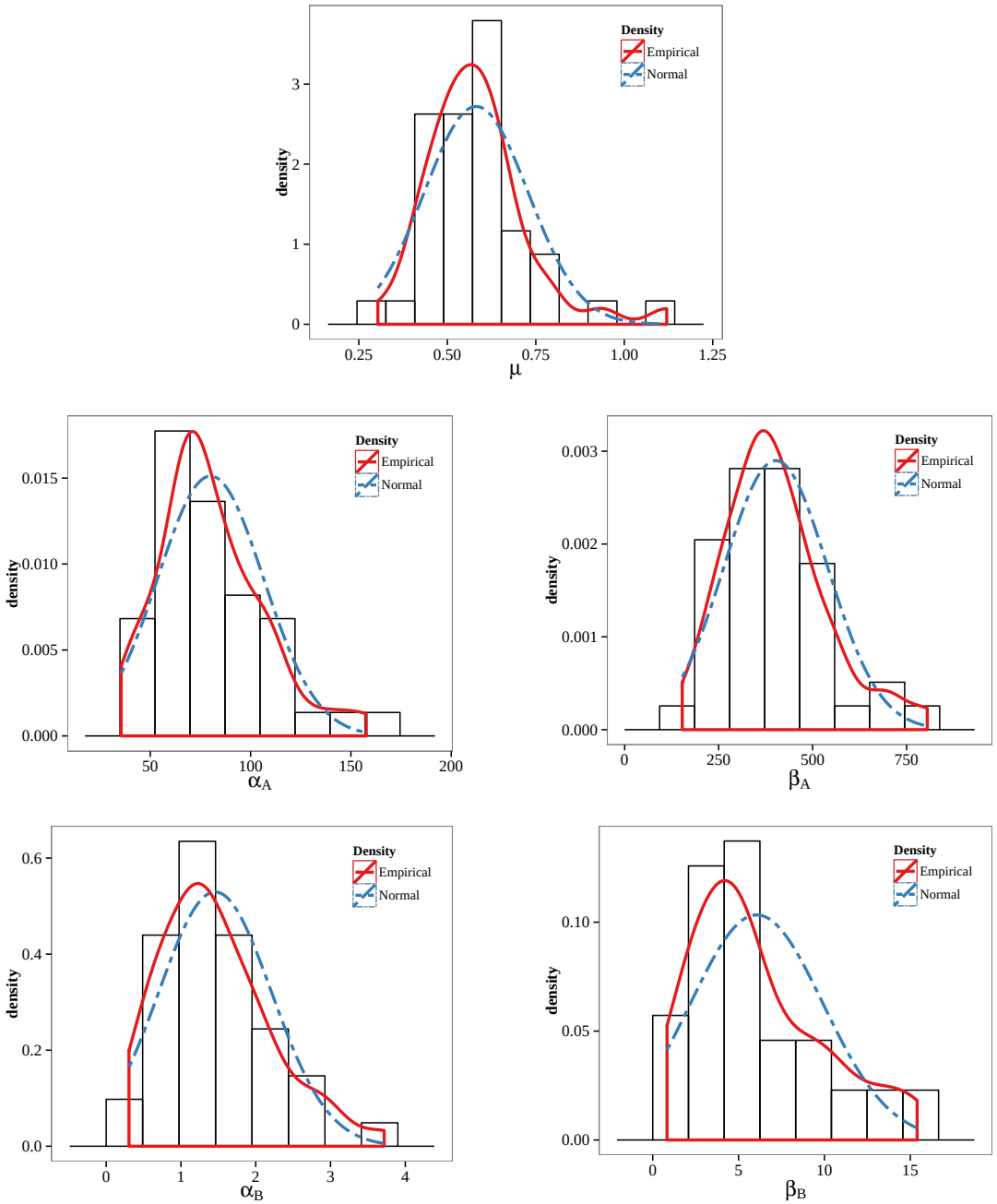

Figure 4.9: *Double Exponential kernel, deseasonalized data*. Distribution of the parameters estimates across the days of the sample. The histogram and the empirical kernel estimates are shown together with the fitted normal density for reference.

11:00-15:00 data				
	μ	H	k	p
Avg. values	0.041	0.051	0.0019	1.04
Std. dev. (%)	36	12	20	4
Avg. uncertainity per day (%)	17	5	17	1.6

Deseasonalized data				
	μ	H	k	p
Avg. values	0.24	0.050	0.00051	1.05
Std. dev. (%)	28	10	24	2
Avg. uncertainity per day (%)	12	4	11	1.0

Table 4.3: *Power law kernel.* Average results across days with corresponding standard deviations. The average uncertainty on the estimates on each day calculated from the Hessian matrix is also given.

4.4.3 Power law kernel

The last model features a power law kernel of the type:

$$w_{\text{PL}}(t) = \frac{H}{(t+k)^p}. \tag{4.8}$$

and the intensity function is:

$$\lambda(t) = \mu + \sum_{t_i<t} \frac{H}{(t-t_i+k)^p}\,, \tag{4.9}$$

where μ, H,k, and p are the constant parameters of the model. As always, μ is the baseline intensity. H is the amplitude of the self-exciting term, p fixes the speed of the decay and must be greater than one for the process to be non-explosive. Finally, k is a short time scale offset to prevent $w_{\text{PL}}(0)$ from diverging. The power law decay does not possess a typical time scale and allows to capture a long range dependence on past events (Bauwens and Hautsch, 2009).

The log-likelihood is:

$$\begin{aligned}\ln \mathcal{L} = -\mu T - \sum_{t_i} \frac{H}{p-1}\left(k^{1-p} - (T-t_i+k)^{1-p}\right) + \\ + \sum_{t_i} \ln\left(\mu + H \cdot R(i)\right),\end{aligned} \tag{4.10}$$

where $R(i) = \sum_{t_j<t_i} w_{\text{PL}}(t_i - t_j)$ and $R(1) = 0$. Note that in this case it is not possible to write a recursion relation as for the exponential kernels, increasing dramatically the computational time required.

Distributions of the results are shown in Figures 4.10 and 4.12. Table 4.3 summarizes the results. We note that the exponent p of the decay is very close to one, which is the onset of non-stationarity. Also, the value of p, with a standard deviation of only $2-4\%$, is quite stable across the days. We note also that for both datasets the offset k assumes values close to the shortest durations, which are of the order of $10^{-3}s$ and 10^{-4} respectively for 11:00-15:00 and deseasonalized data.

In Figures 4.11 and 4.13 a graphical evaluation of the fitting capabilities of this model is presented through the usual QQ-plots. We refer to the next section for a more complete analysis of the performance.

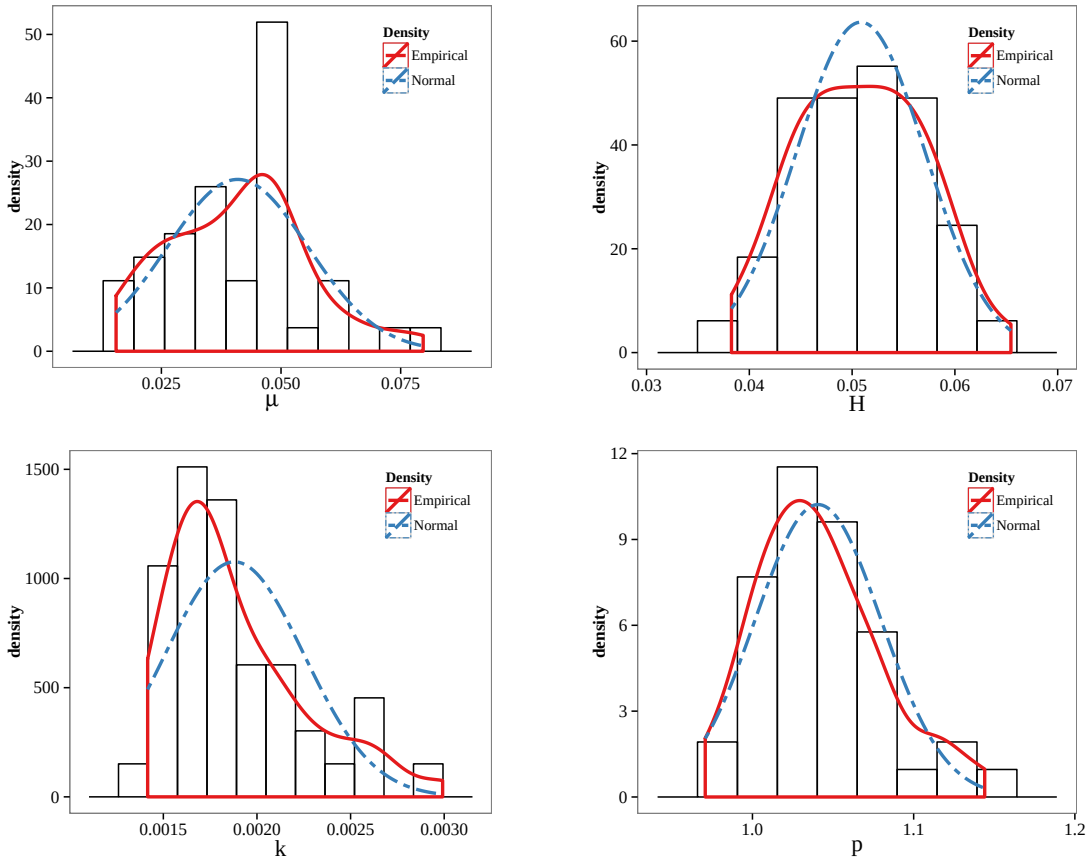

Figure 4.10: *Power-law kernel, 1100-1500 data.* Distribution of the parameters estimates across the days of the sample. The histogram and the empirical kernel estimates are shown together with the fitted normal density for reference.

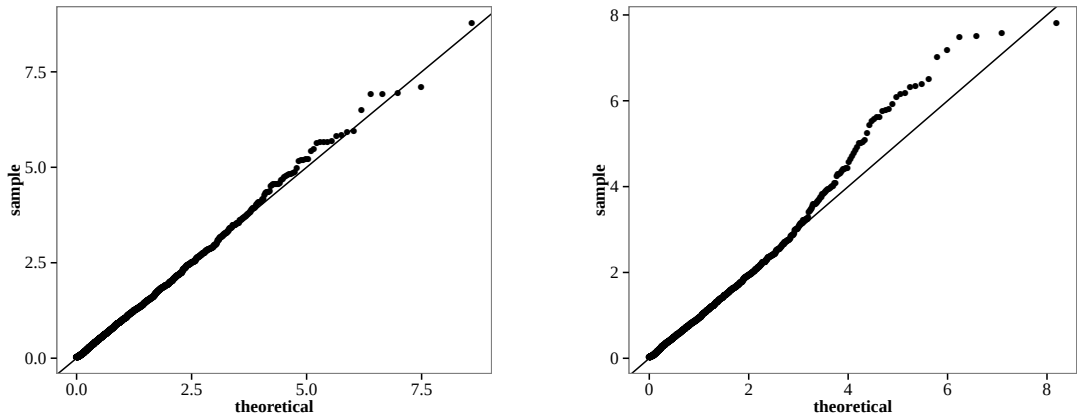

Figure 4.11: *Power law kernel, 11:00-15:00 data.* QQ-plot of residuals against standard exponential distribution for two days obtained with the power law kernel. The graph on the left shows one among the best results, that on the right one among the worst.

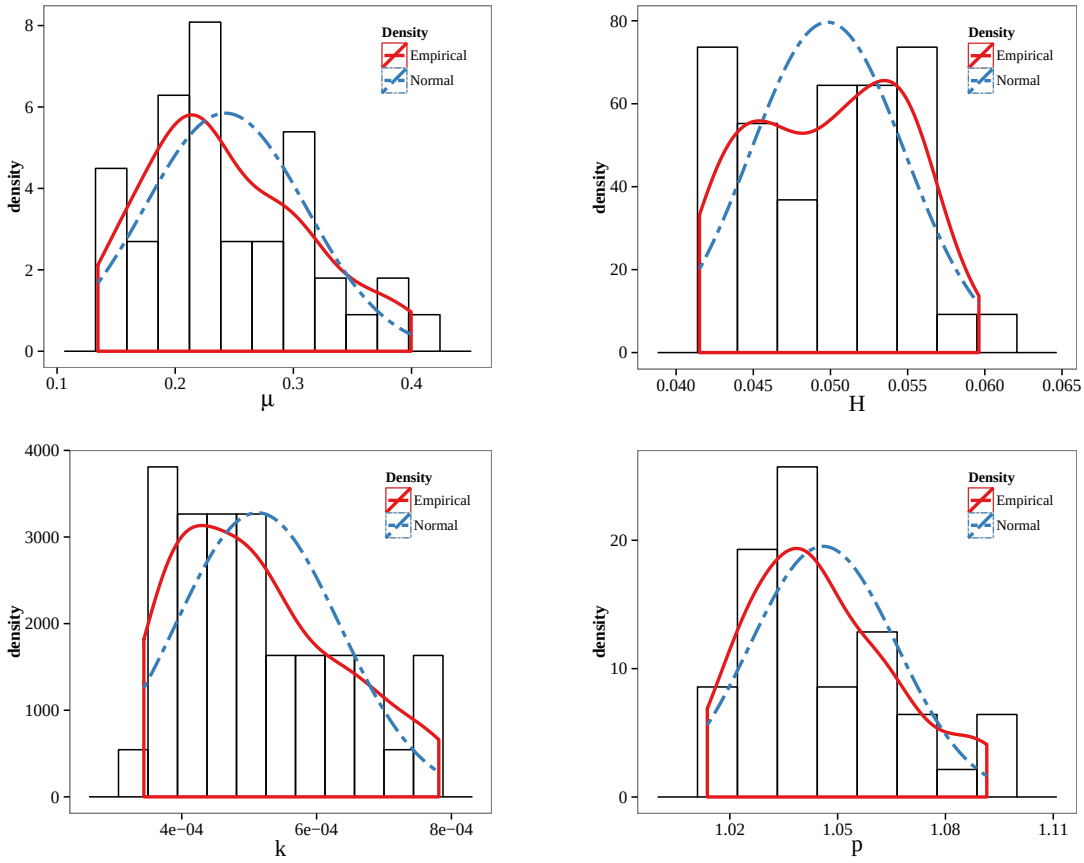

Figure 4.12: *Power-law kernel, deseasonalized data.* Distribution of the parameters estimates across the days of the sample. The histogram and the empirical kernel estimates are shown together with the fitted normal density for reference.

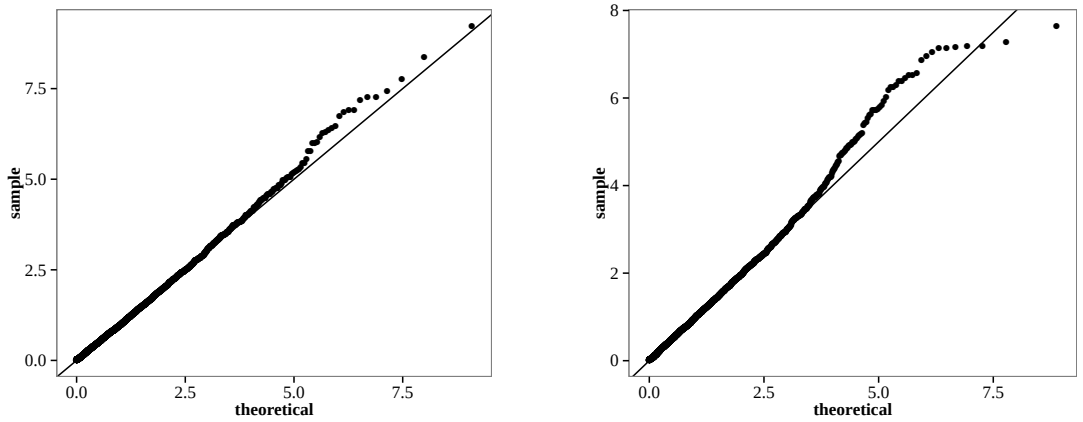

Figure 4.13: *Power-law kernel, deseasonalized data.* QQ-plot of residuals against standard exponential distribution for two days obtained with the power law kernel on deseasonalized data. The graph on the left shows one among the best results, that on the right one among the worst.

4.5 Test and discussion

As pointed out in Subsection 2.3.3, the starting point for goodness-of-fit assessment is the analysis of the residuals $\Lambda_i = \int_{t_i}^{t_{i+1}} \lambda(t)\mathrm{d}t$. If the model perfectly describes data these should be i.i.d. exponential random variables.

In Tables 4.4 and 4.5 the results of the ED test defined in (2.45) for the two dataset are summarized: the number of days for which the test rejected the null hypothesis of exponentially distributed residuals is reported for each type of kernel. In the same tables the outcomes of a Box-Ljung test that the first 15 autocorrelations are zero are also shown. The significance level was set in both cases at 5%.

It was already clear from the graph of the previous sections that the single exponential kernel is not capable of a satisfactory description of the data, and the tests confirms this, rejecting the null for all the days in the sample. The double exponential and power law kernel provide better alternatives. Between the two, both the graphical analysis and the ED test suggest that the power law kernel gives the best fit. Moreover, according to the ED test, only the power law kernel achieves consistent results on both dataset. In fact the DE kernel fails to pass the ED test on the deseasonalized data, although graphical analysis is not as severe. The Box-Ljung test results again favour the power law kernel, at least for the 11:00-15:00 dataset. However, on most days some correlation between the residuals is detected for all the models and on the deseasonalized sample the null is rejected for almost all days.

To compare further double exponential and power law kernels performances, Akaike Information Criterion (AIC) (Akaike, 1974) was also employed. This model selection test is similar to a likelihood test, but takes into account the different number of parameters. It is used to decide which, in a set of models, has the highest relative probability of minimizing the information loss [1], and hence describes better the data. AIC score is defined as:

$$\mathrm{AIC} = 2k - 2\ln\mathcal{L}, \tag{4.11}$$

where k is the number of the model's parameters and $\mathcal{L}$ is the likelihood at the optimal values. Again the power law kernel was consistently better among all days. Finally, the superiority of the power law kernel can be appreciated by noting that the estimated parameters are very similar across all days of the sample, while those relative to the double exponential kernel have wider fluctuations. However, the very good fit provided by the power law kernel is somewhat counterbalanced by its very high computational cost compared to the exponential family.

	ED test			*Box-Ljung test*		
	SE	DE	PL	SE	DE	PL
Rejected	42	35	14	42	39	33
Not rejected	0	7	28	0	3	9
% rejected	100.00	83.33	33.33	100.00	92.86	78.57
% not rejected	0.00	16.67	66.67	0.00	7.14	21.43

Table 4.4: *11:00-15:00 data.* Summary table of excess dispersion and Box-Ljung test. ED test is a test of exponentiality, Box-Ljung tests the null hypothesis that the first 15 autocorrelation are zero. For each type of kernel, the number of days for which the test rejects (at 5% level) is reported.

Another comparison of the different kernels is achieved by looking at the duration distribution produced by simulation with optimal parameters. For each day we simulate a Hawkes process with the three different kernels presented, using the best parameter estimates for that day. Then, we compute the durations from each simulated process and from the original data. Finally, we pool the computed durations of each type from all days in a single sample. In the end, we have four samples of durations, one for the single exponential model, one for the double exponential model, one for the power law model, and one for the

[1] The information loss is the information we lose in using a certain model to describe reality.

	ED test			*Box-Ljung test*		
	SE	DE	PL			
Rejected	42	42	15	42	41	41
Not rejected	0	0	27	0	1	1
% rejected	100.00	100.00	35.71	100.00	97.62	97.62
% not rejected	0.00	0.00	64.29	100.00	2.38	2.38

Table 4.5: *Deseasonalized data.* Summary table of excess dispersion and Box-Ljung test. ED test is a test of exponentiality, Box-Ljung tests the null hypothesis that the first 15 autocorrelation are zero. For each type of kernel, the number of days for which the test rejects (at 5% level) is reported.

11:00-15:00 data			
	$\mu_{\mathbf{SE}}$	$\mu_{\mathbf{DE}}$	$\mu_{\mathbf{PL}}$
Avg. value (s^{-1})	0.120	0.08	0.041
Correlation with activity	0.97	0.58	0.24
Deseasonalized data			
	$\mu_{\mathbf{SE}}$	$\mu_{\mathbf{DE}}$	$\mu_{\mathbf{PL}}$
Avg. value	0.79	0.58	0.24
Correlation with activity	0.99	0.81	0.40

Table 4.6: Values of the baseline intensity μ for all the models along with their correlation with trading activity.

real durations. We report here only the results for the 11:00-15:00 dataset, both because the fit on these data is better and because there are no interference from the time transformation procedure. Since the original data have a finite time resolution, the simulated durations are rounded to the millisecond, and those shorter than one millisecond are set equal to one millisecond.

In Figure 4.14 the QQ-plot of simulated durations against real durations are shown together with a comparison of the corresponding empirical cumulative distributions. Again we note that the power law kernel gives the duration distribution which is closest to the true one. To give a complete picture, in Figure 4.15 we show also the probability density estimates for the four distributions. To improve readability we divided the range spanned by the durations in three intervals and the densities are compared in each of them.

We observe that all the three models underestimate the frequency of very short durations, albeit the power law kernel gets closer to the original data. We note also that the single exponential kernel, which has a single time scale, overestimates the number of durations of the order of its characteristic time ($\approx 0.03\,s$), while it produces too few shorter and longer durations. The double exponential kernel suffers from the same problem in correspondence of the shorter time scale ($\approx 0.01\,s$) which has the highest weight. However its performance is already much better than the one of the SE kernel. This suggests that the original process needs many time scales to be correctly described.

A final study of the relative performance of the three models concerns the baseline intensity μ. This can be thought of as the part of the intensity not explained by the self exciting mechanism. It is the rate of *immigrants* in the immigrants-offspring representation of a linear Hawkes process presented in Section 2.3. In Table 4.6 we report again the values of μ obtained with the three kernels together with their correlation with the average trading activity, measured for each day as the number of trades per unit time.

From the table it emerges that for the single exponential model the mean value of μ is the highest and μ itself is almost perfectly correlated with trading activity. Hence, in this case a lot of the activity is attributed to the baseline intensity and this is in line with the poor performance of the model. The situation

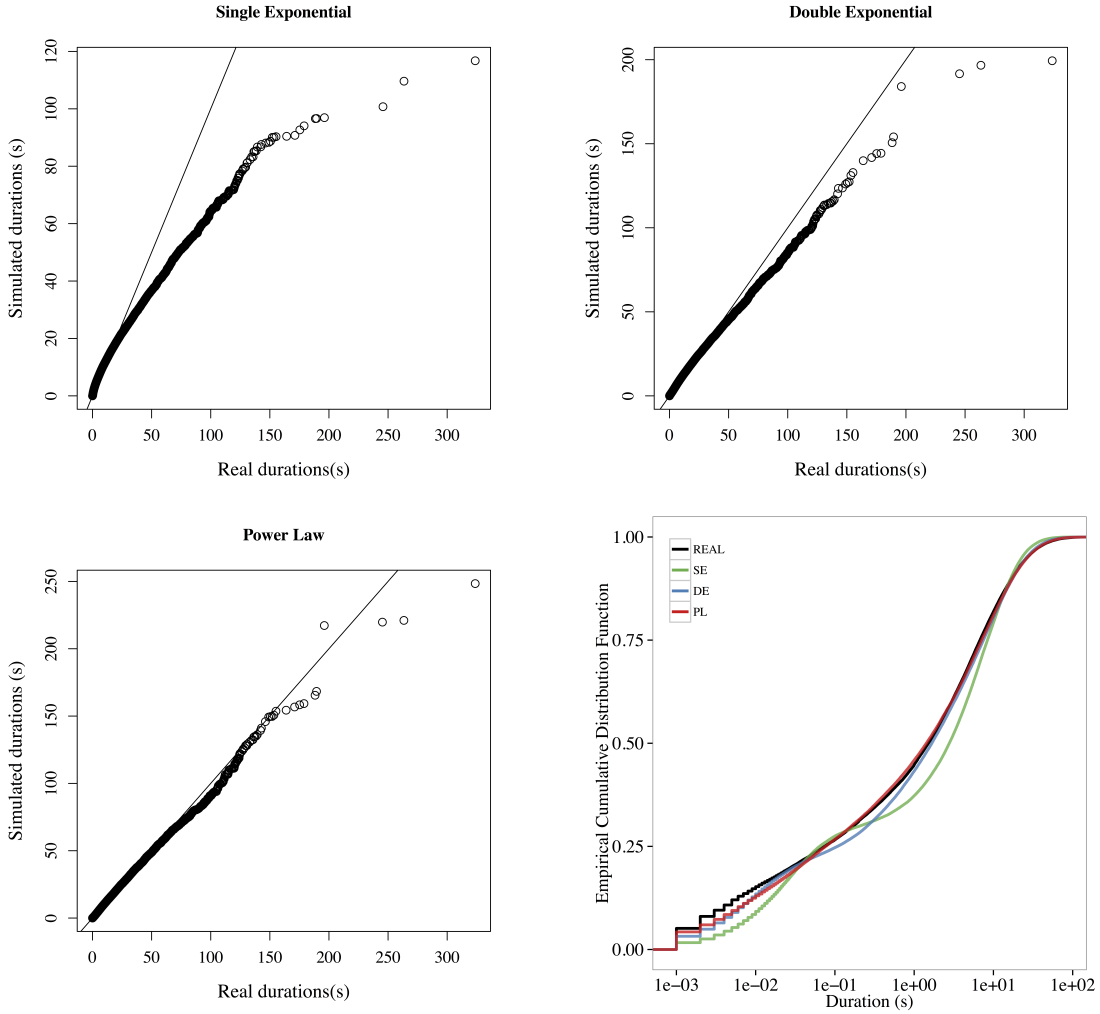

Figure 4.14: *11:00-15:00 data*. QQ-plots of simulated durations against real durations for the three kernels. The plot on the bottom right is the Empirical Cumulative Distribution Function for the simulated as well as the original durations.

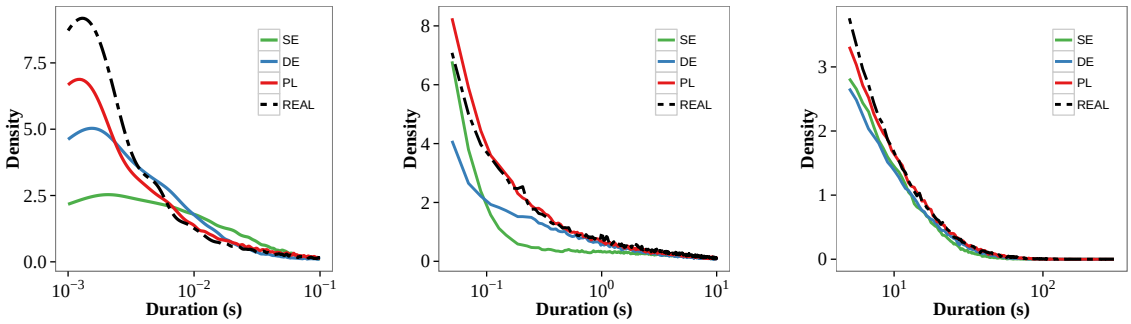

Figure 4.15: *11:00-15:00 data*. Probability density estimates of the duration distributions. Three range of durations are shown separately for clarity, from left to right: short durations, medium durations and long durations.

improves greatly with the double exponential kernel and especially with the power law one. For the 11:00-15:00 data, the correlation of μ_{PL} with activity is even not significant at 5% level, i.e. we cannot reject the null of μ and trading activity being uncorrelated.

As a final remark, by looking at the values obtained for the exponent of the power law kernel, we note that the results are very close to one, which is the limit for a stationary process. Hence, the best fit to the data seems to be achieved with a slowly decaying kernel. This is in agreement with the findings of Bacry et al. (2012) on Bund futures data, who, using a non-parametric approach, report a power law kernel with exponent close to unity. Also Hardiman et al. (2013) reports similar results on equity data. The model main limit seems to resides in the non completely satisfactory reproduction of the duration distribution for very short lags.

Chapter 5

Modelling FOREX data with Hawkes processes

In this chapter we report the results of the application of univariate Hawkes models to FOREX data. Section 5.1 describes the dataset we used and the transformations we made before using data. Section 5.2 gives the details of the model employed. In Section 5.4 we discuss overall model performance, while in Section 5.5 we examine the meaning of the values we found for the parameters an their relation with trading activity. All the results discussed in this chapter are original.

5.1 Data

The dataset is the one described in Section 3.1.2. Here the events are changes in either the best bid or the best ask price. We chose this type of event for two reasons. First, there are far more quote-change events than trade events, hence estimation can benefit from larger sample size. The second reason is due to the nature of the market we are studying. Interbank platforms like EBS are used by market makers for two main purposes:

- As the starting point for setting prices offered to clients. That is the bid and ask price proposed to clients are fixed starting from those available on EBS.
- To rebalance position. In trading with clients banks try to avoid taking large positions in any currency in either direction. In other words, if a clients sells them a large quantity of a currency, they try to re-sell this quantity to other clients earning the spread. If banks are not able to reblance their position internally. i.e. with their clients, then they use the interbank platform to do so, trading with other banks.

So we can say that the quote process on EBS influences the whole FOREX market much more than the trade quote does. This is the other reason why we opted for modelling the quote process.

As anticipated in Section 3.2.2 we decided to use only the data relative to the London working hours, namely from 07:30 to 16:30 London time. This is done to avoid mixing time intervals with very different activity. Also, as we showed, market activity on FOREX is concentrated in London working hours. In the following we will refer to this specific time window as a "day".

For this dataset we have to deal with two properties of the data that may influence the model, namely intraday pattern and the influence of the snapshot time scale.

Regarding intraday seasonality, during the London working hours (LWH) window, we note from Figure 5.1 that the ratio between the peak of activity around 15:30 and the bottom at midday is ≈ 2. This is much less than in the equity case, where the figure was about 7. This, together with the much higher number of events per unit time, suggests that it may be possible to use the whole period without transforming

the time. To test this hypothesis we estimated the parameters of a univariate Hawkes model with double

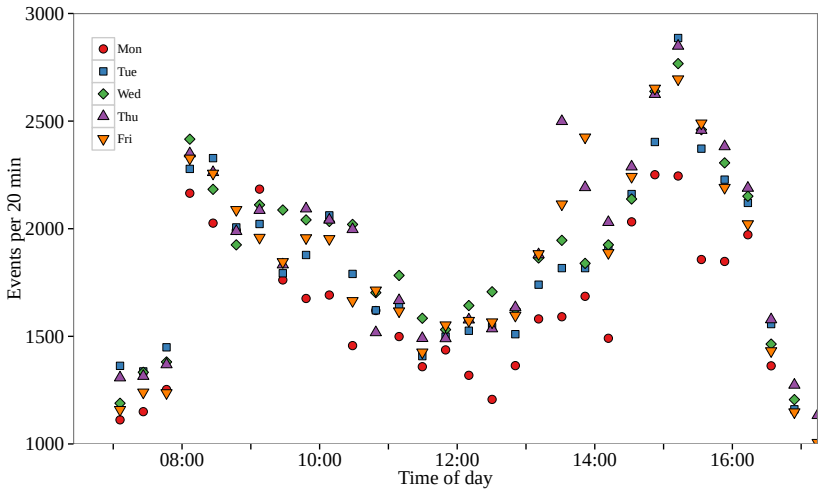

Figure 5.1: Activity pattern during London working hours for EUR/USD measured in number of events (quote changes) per 20 min. Different days of week are shown.

exponential kernel as in Section 4.5 on four separate periods inside the LWH window. The intervals are chosen to be:

07:30-09:00
09:00-11:30
11:30-14:00
14:00-16:30

The estimation was performed on 40 days (February and March) of the EUR/USD sample. The time series of the estimated parameters is shown in Figure 5.2, we note that the estimate for α_A and β_A are consistently lower in the 14:00-16:30 window, while values for the other windows are closer one with each other. Differences in the other parameters are less evident though still noticeable.

In the end we decided to test both deseasonalized and untransformed data. Working with original data is appealing because the parameters have a clear temporal meaning. Also, with applications in mind, it is desirable to have models directly applicable to data.

The other issue that emerged in section 3.2.2 is the clustering of the durations around multiples of 100 ms. This is an artefact due to the temporal resolution EBS employs in sampling the order book and is not related to the "true" event arrival process, i.e. the quote change. To partially remove this effect we adopt a randomization procedure:

1. We start from *ebsReferenceTime*, for which the minimal separation of 100 ms is strictly valid.
2. We subtract from each time stamp expressed in second a random number uniformly distributed in (0,0.1]

The rationale behind this approach is that when a snapshot is recorded, we know that at least one event took place in the previous 100 ms. This event could have happened at any instant in the 100 ms before the snapshot was recorded and hence, in absence of other information, we assign it to a random instant within the 100 ms interval. Clearly, this is only an approximation of the true quote-change process. In fact, as we noted in Section 3.1.2, no information is kept on how many quote events occurred in the 100 ms interval. The effects of this procedure on the duration distribution are showed in Figure 5.3.

Combining the pattern-removing procedure and the randomization procedure we obtained three samples for each pair on which we test the model:

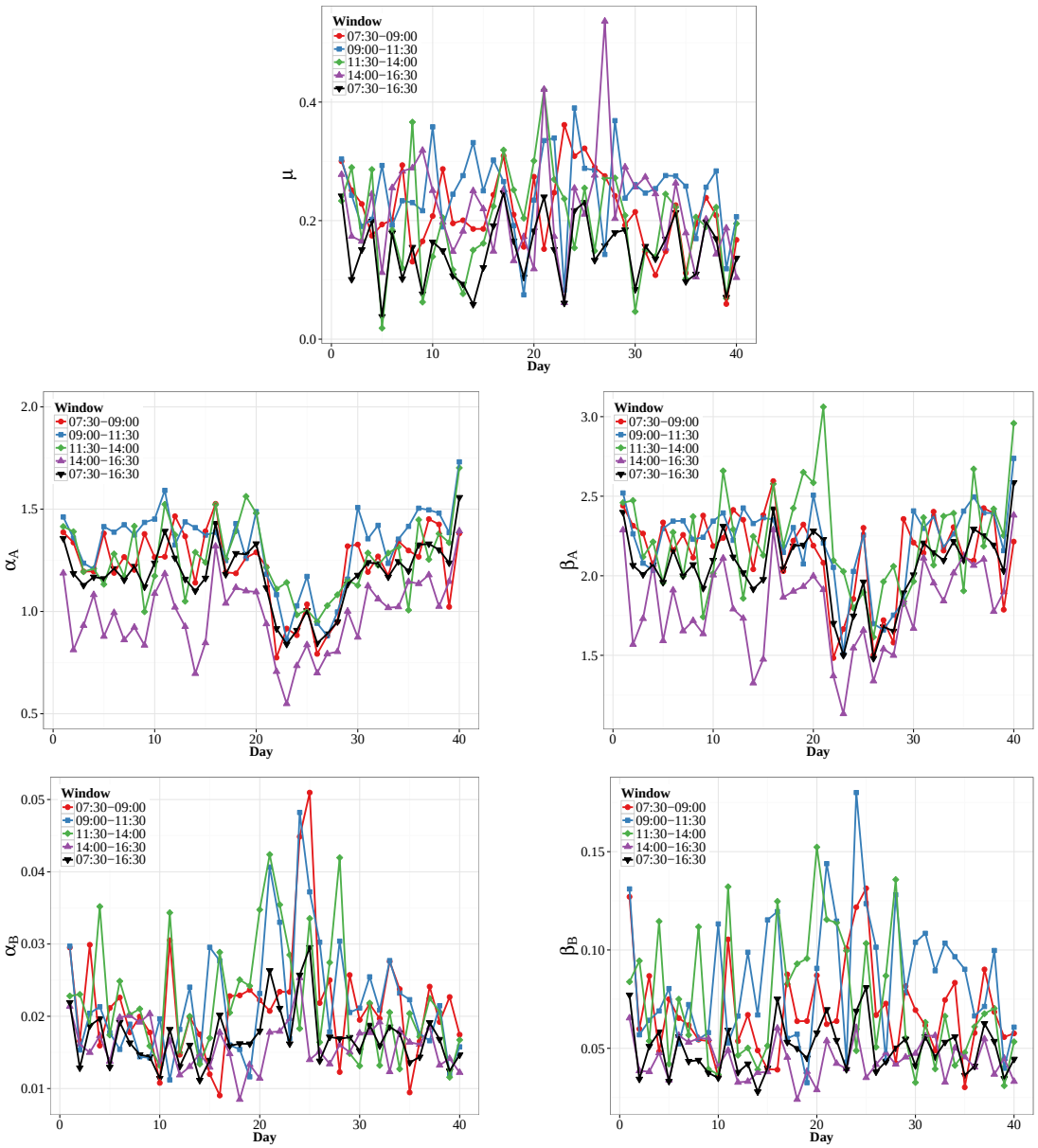

Figure 5.2: Time series of the estimated parameters over 40 days of February and March for different time windows.

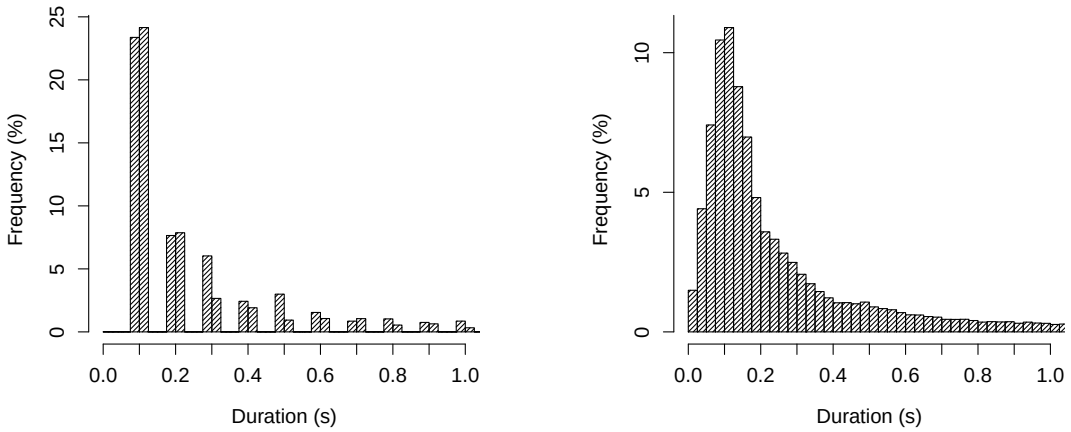

Figure 5.3: Effects of the randomization procedure for a typical day. Left: duration distribution zoomed on [0,1] before the randomization. Right: same day after randomization.

- *Original sample*: original series of ebsMarketUpdate times restricted to LWH.
- *Randomized sample*: randomized times obtained from ebsReferenceTime.
- *Randomized and deseasonalized sample*: obtained from the randomized sample with the same procedure used on equity data.

5.2 Model specifications

Given the results of Chapter 4, we decided to use only the double exponential and the power law kernels, which demonstrated good performances. However, for computational reasons we do not use a true power law kernel but instead we use an exponential sum that carefully approximates a function with a power law tail. In particular we adopt the same specification of Hardiman et al. (2013), which reads

$$w_{\mathbf{PL}}(t) = \frac{n}{Z}\{\sum_{k=0}^{M-1}(a_k)^{-p}e^{-\frac{t}{a_k}} - Se^{-\frac{t}{a_{-1}}}\} \tag{5.1}$$

where

$$a_k = \tau_0 m^k. \tag{5.2}$$

The parameters are n, p and τ_0. M is the number of exponential terms and m is a scale parameter. We fix $M = 15$ and $m = 5$ as in Hardiman et al. (2013). The parameter m controls how well a single exponential approximates the power law in a decade. If one wishes the approximation to be valid over many orders of magnitude, high values of m and M are necessary. Simply increasing m extends the range on which the approximation holds, but produces oscillations, as illustrated in Figure 5.4. Z and S are constants such that $w_{\mathbf{PL}}(0) = 0$ and $\int_0^\infty w_{\mathbf{PL}}(t)\mathrm{d}t = n$. Thus, the condition for stationarity is $n < 1$. The main difference with the true power law kernel used in the equity section is the presence of a smooth cut-off at short lags, provided by the negative exponential, whose time scale is regulated by τ_0 once m is fixed. The two kernels are different also at very large values of t, since the approximated power law kernel presents an exponential cut-off at large t. This difference is however irrelevant, in fact, for our choice of M and m, it manifests itself only for t larger than $\approx 10^9$s, that compared to the duration of trading day is the same as infinity. This approximation of the power law in terms of an exponential sum allows the log-likelihood to be computed recursively, reducing computational cost form $O(N^2)$ to $O(N)$, N being the sample size. This is particularly relevant given the large number of data available.

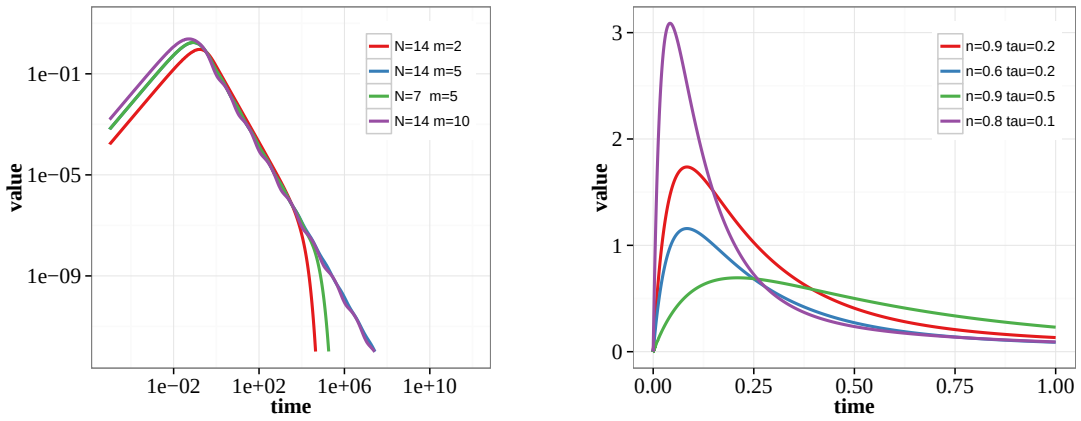

Figure 5.4: Approximated power law kernel. Left: effects of the changes in m and M, $n = 0.9$, $\tau_0 = 0.2$,$p = 1.5$. Right: shape of the kernel for different combinations of n and τ_0 at small values of t, $p = 1.5$.

5.3 General procedure

We adopt a similar procedure to the one described in Section 4.3. We obtained the best parameters estimates for each day maximizing log-likelihood. Here we use R optimization routine *nlminb* with a set of 10 different starting point.

We repeat the procedure for each pair of currencies and for both types of kernel. Uncertainties on the best parameters estimates are calculated using the inverse of the Hessian matrix at the maximum, as discussed in Subsection 2.3.2.

5.4 Overall model performance

The first point we address is the performance of the model in relation to the dataset. In other terms, we assess the impact of seasonality and duration clustering on goodness-of-fit. To this end we proceeded as follows. For each pair of currencies:

1. We estimate the best parameters day by day on the three different samples (original, randomized and randomized with pattern removal).
2. We compute the residuals using the best estimates for each day, for all the three data samples.
3. We aggregate the residuals from different days obtaining three vectors of residuals, one for each sample of data.
4. We compare the distributions of the residuals with the standard exponential distribution.

The results using the double exponential and power law model are shown in qq-plots of Figure 5.5 for USD/JPY. The behaviour for the other pairs is qualitatively the same. We observe that the influence of the data type on the right tail of the distribution is negligible, though the models perform slightly better on original and randomized data. Instead, from the log-log plot, we see that the left tail of the distribution is quite different for randomized and non-randomized data. Here, the fit on the randomized data is much better, albeit not perfect. Also, no sensible difference can be observed between randomized and deseasonalized data, and simple randomized data. Figure 5.6 shows the histograms of the residuals obtained from original and randomized data. This figure highlights the problem in the first part of the distribution. This problem is somewhat expected for the test on original data since we are fitting a Hawkes process, which produce a continuous distribution of durations, on data that present an (almost) discrete duration distribution. The discrete nature of durations in original data is clearly more evident at short

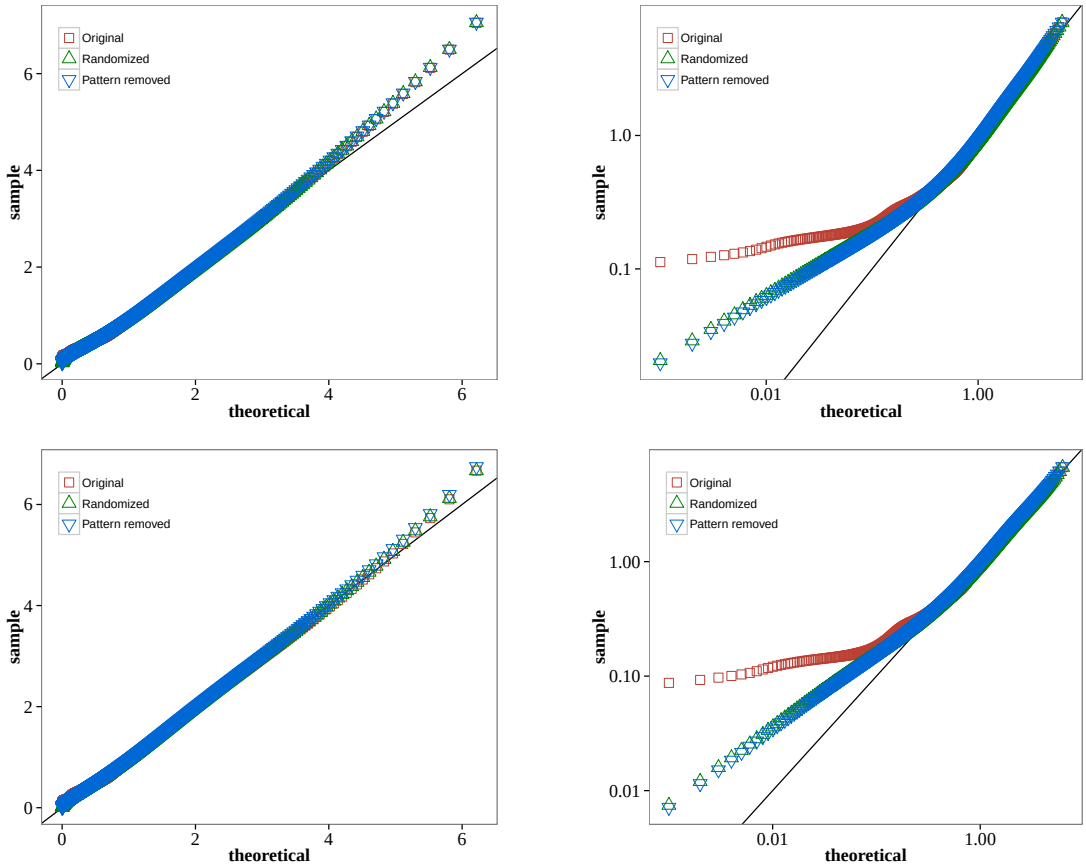

Figure 5.5: QQ-plots of residuals against standard exponential quantiles for the three data type. Top: double exponential kernel, same data in linear (left) and log-log (right) scale. Bottom: power law kernel, linear scale (left) log-log scale (right).

time scales, where durations are comparable with the snapshots time resolution. As it turns out, in fact, residuals with values in $(0, 0.6]$, that are those where deviations from exponential distribution are more severe, correspond to durations around 100 ms. Randomization does not resolve completely the issue. Here the problem is that the kernels we use are not able to produce a duration distribution with enough weight around 100 ms as is observed both in original and randomized data. In Figure 5.7 we compare the durations distributions from simulations with that of randomized data. The simulations produce too many short durations. The power law kernel, which is equipped with an exponential cutoff at short lags performs better, while the double exponential kernel that present its maximum at $t = 0$ has the worst result. A shape of the kernel more sharply peaked around 100 ms with a slowly decaying tail could perform better, however it is not trivial to combine this requirement with a low number of parameters and also a form that retain the possibility of iterative computation of likelihood. A non parametric approach for the estimation of the best form of the kernel like the one of Bacry et al. (2012) would be also interesting to shed light on the true optimal form for the kernel.

From this analysis we can conclude that:

1. The clustering of the durations around 100 ms observed in the original data does impact model performance.

2. Randomizing the times sensibly improves the model fit, albeit problems still remain as the model cannot reproduce correctly the duration distribution at small values.

3. Finally, removing the pattern does not seem to further improve results.

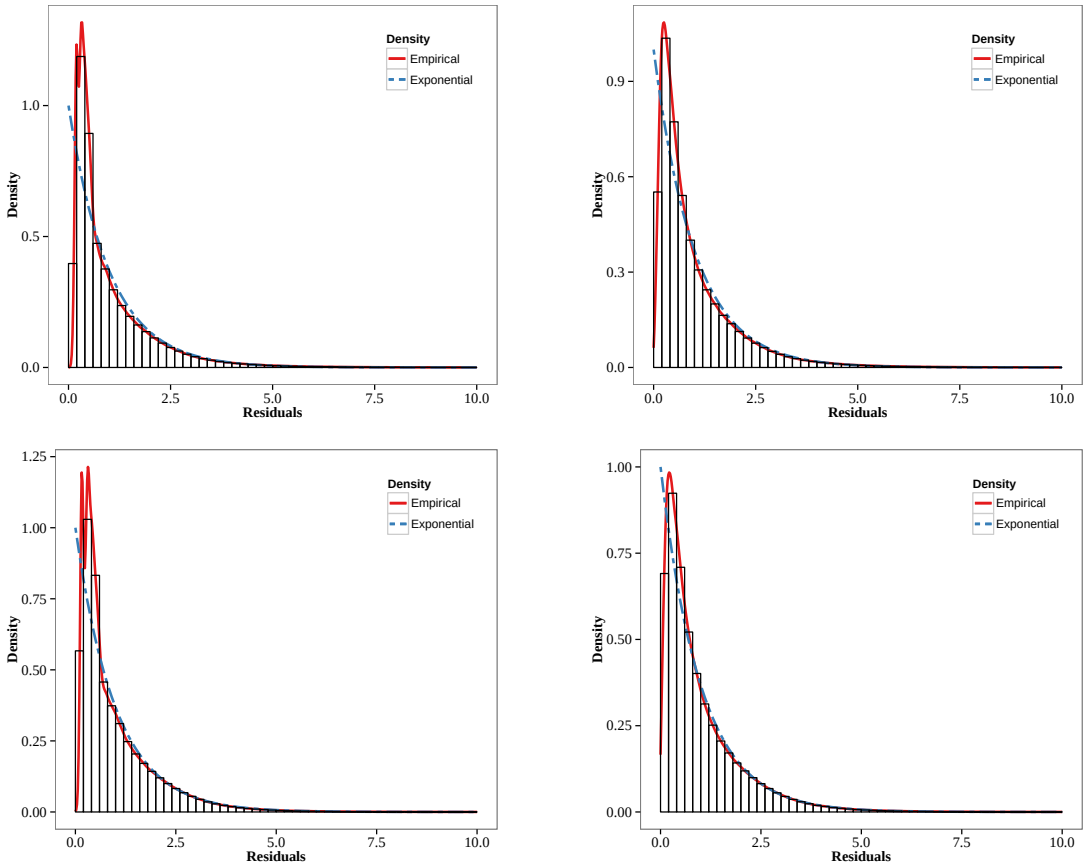

Figure 5.6: Histograms of the residual distribution, with estimated empirical density (red) and the reference exponential distribution (blue). Top: double exponential kernel, original data (left) randomized data (right). Bottom: power law kernel, original data (left), randomized data (right). The bin size is 0.2

Hence, we decided to limit the analysis presented in the following to the randomized sample, with potential references to the original data where pertinent.

Regarding the comparison of the two kernels, we have already seen from the residuals plots that the power law kernel appears to perform better. This is confirmed also by likelihood values, which are consistently higher for the power law kernel.

We do not repeat here the analysis of ED test. In fact, the problem with short durations that affects both kernels leads to the rejection of the hypothesis of exponentially distributed residuals.

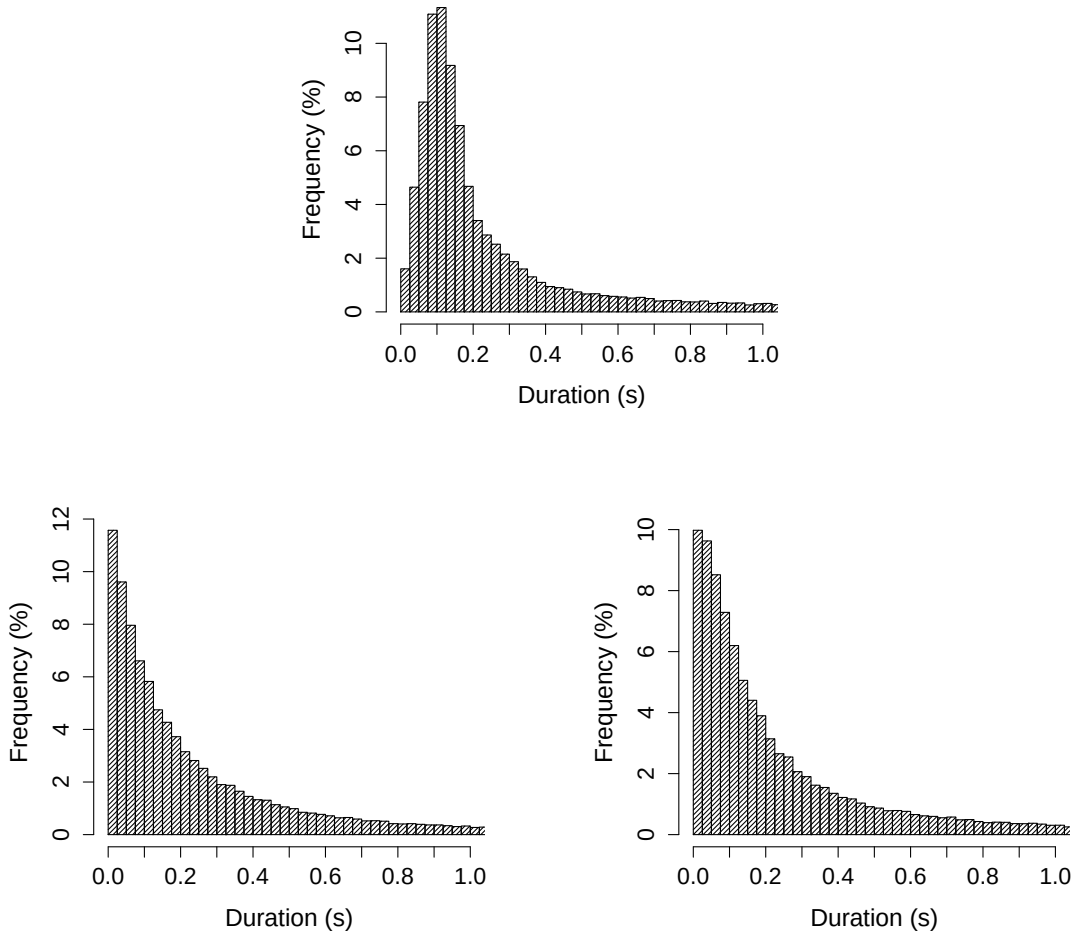

Figure 5.7: Real and simulated durations for a particular day of our sample (EUR/USD day 61). Real durations are shown in the top figure, durations distribution of a process simulated with double exponential kernel are reported on the bottom left, while on bottom right we show the durations distribution from data simulated with power law kernel. The bin size used in the histograms is $0.025s$. The x axis is limited to 1 to appreciate the differences at short durations.

5.5 Discussion of parameters values

After having discussed the overall results and limits of the model, we now turn the analysis on the values obtained for the parameters. We will focus on results from the power law kernel. In Figure 5.8 we report the time series of parameters estimated on EUR/USD data with the power law kernel. Trading activity measured in events per second is also shown. First of all we note that the periods before and after the tick increase, signalled by the dotted line, are clearly distinguishable. This is true also for the other pairs. After the tick increase the number of quote changes per second drops sharply as expected. In fact, after the increase a change in a quote implies a larger movement in the posted price. This has consequences also on the durations distribution. As it is visible in Figure 5.9, the weight of longer durations increases after the tick change.

In Table 5.1 we report the average value of the parameters obtained with the power law kernel before and after the tick increase for all the three pairs. The uncertainties on the values as estimated from the Hessian of the log-likelihood are about 10% on μ, 2% on n and τ_0 and about 1% on p. μ is sensitive to the overall activity and hence it is not surprising that it follows its trend.

Also p varies sharply after the tick increase. In particular, as p controls the power law decay, the kernel is

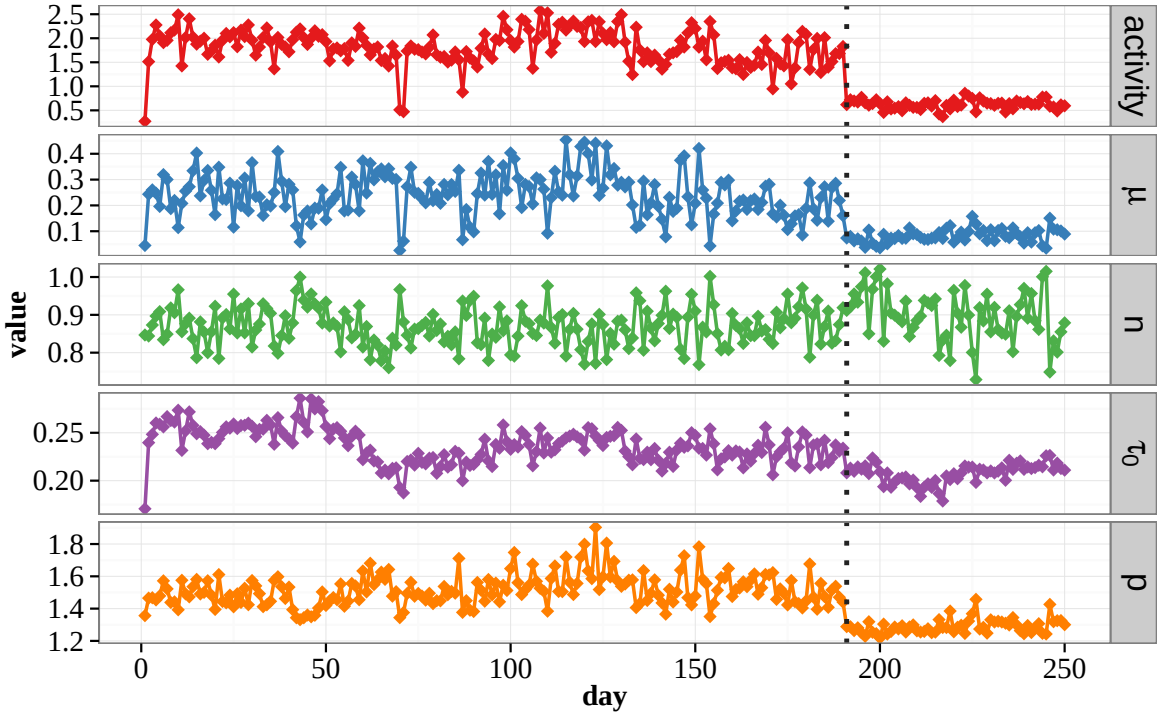

Figure 5.8: Time series of the parameters estimates for EUR/USD. Market activity in events per second is also shown. The dotted line signals the tick size increase.

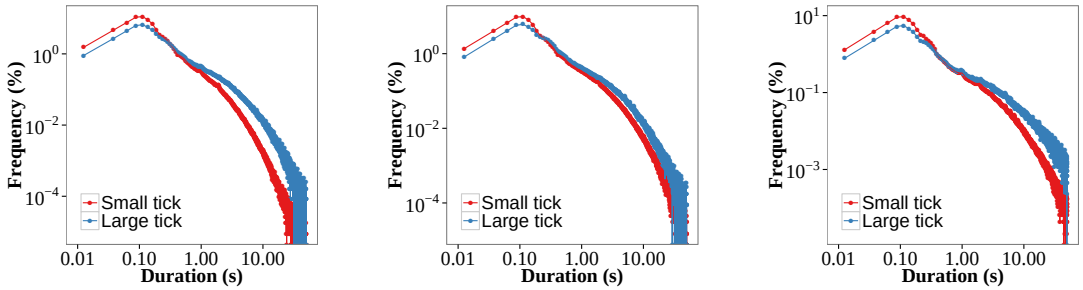

Figure 5.9: Durations distribution for the three currency pairs before and after the tick increase. The scale is logarithmic on both axes. To create the graphs, durations were binned using $0.025s$ long intervals.

found to decay more slowly after the change in the ticksize. This in turn enhances the number of long durations produced by the model, in accordance with what is observed on the data.

The parameter n increases for all pairs after the change in the tick size. In the representation of linear Hawkes processes as made of immigrants and offspring points, n is directly connected to the fraction of events generated endogenously. In particular, it is the expected number of additional points that any single events generates. Moreover, if we assume the process is stationary, n and μ are related through the relation $\overline{\lambda} = \mu/(1-n)$. So n is the fraction of the average rate explained by the self-exciting mechanism.

Then higher values of n can be interpreted in two ways. If we suppose the model is an equally valid representation of the observed process before and after the tick increase, we can conclude that after the tick increase the self exciting mechanism is stronger. On the other hand, the durations distribution after the increase in the tick is less peaked at 100 ms and we have already noted that this feature of the data is not well captured by the model. This fact weight on goodness-of-fit. As it is possible to note from figure 5.10, residuals from the large tick data are closer to the exponential distribution. Then, an alternative interpretation of the increase of n after the tick increase is that the self-excitation of the true process is approximately the same before and after the tick increase and the observed increase in n is a result of the

	Before			After		
	EUR/USD	EUR/JPY	USD/JPY	EUR/USD	EUR/JPY	USD/JPY
activity (events/s)	1.82 (20)	1.08 (28)	0.76 (31)	0.62 (15)	0.59 (18)	0.22 (26)
μ (s^{-1})	0.24 (35)	0.17 (44)	0.12 (36)	0.08 (31)	0.08 (32)	0.03 (36)
n	0.87 (6)	0.86 (9)	0.84 (6)	0.90 (7)	0.91 (7)	0.95 (8)
τ_0 (s^{-1})	0.24 (8)	0.20 (11)	0.21 (7)	0.21 (4)	0.22 (16)	0.18 (4)
p	1.51 (6)	1.46 (10)	1.52 (7)	1.29 (3)	1.26 (3)	1.22 (3)

Table 5.1: Average values of the parameters before and after the tick size increase. Standard deviations across the sample expressed in percent is also reported in parentheses.

better fit of the model on data.

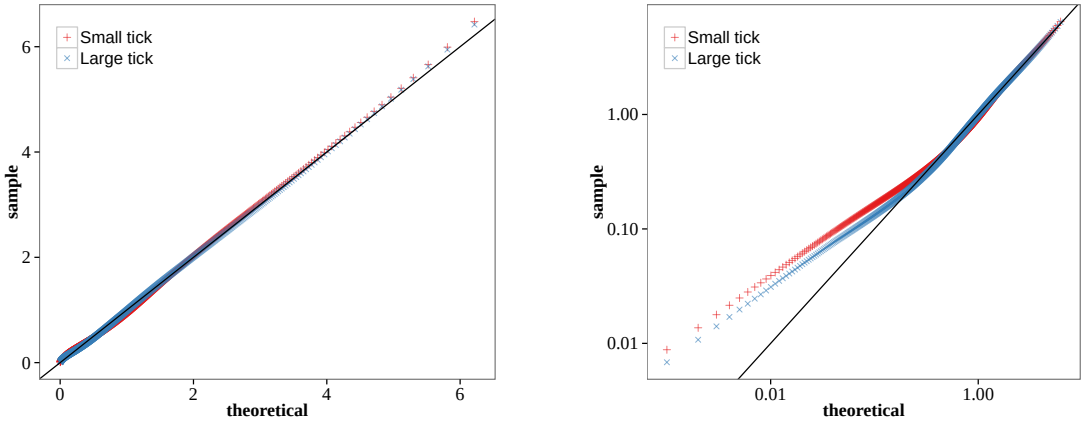

Figure 5.10: QQ-plots of residuals against exponential distribution for EUR/JPY, before and after the tick increase. Left graph is in linear scale, right one in log-log scale.

Finally we estimated the best parameters considering the different days as independent realizations of the same process. Thus, we maximized with respect to the parameters the total log-likelihood given by the sum of the log-likelihoods of the various days. We considered separately the periods before and after the tick size increase. The results are reported in Table 5.2. The results are in line with those presented in Table 5.1. We note, however, that we obtained higher values of n and consequently also lower values of μ.

	Small tick			Large tick		
	EUR/USD	EUR/JPY	USD/JPY	EUR/USD	EUR/JPY	USD/JPY
activity (events/s)	1.82 (20)	1.08 (28)	0.76 (31)	0.62 (15)	0.59 (18)	0.22 (26)
μ (s^{-1})	0.147 (0.6)	0.115 (0.6)	0.085 (0.5)	0.068 (2)	0.059 (2)	0.0212 (2)
n	0.926 (0.07)	0.906 (0.09)	0.893 (0.08)	0.932 (0.3)	0.956 (0.3)	0.987 (0.4)
τ_0 (s^{-1})	0.232 (0.09)	0.191 (0.10)	0.203 (0.10)	0.2059 (0.3)	0.2155 (0.3)	0.1792 (0.4)
p	1.430 (0.05)	1.381 (0.06)	1.445 (0.06)	1.266 (0.1)	1.237 (0.1)	1.203 (0.1)

Table 5.2: Parameters obtained maximizing total log-likelihood before and after the tick increase. In parentheses, for the parameters estimates we report the uncertainties calculated from the Hessian matrix, values are in percent. For activity standard deviation is shown.

Overall, the fit provided by the Hawkes model with the power law kernel is pretty good. In line with Hardiman et al. (2013), we find that a large part of the activity is explained by the self exciting mechanism (n close to 1) and hence appears to be endogenously generated.

Chapter 6

Hawkes process description of exogenous contributions to market activity

In this chapter we present a model based on the Hawkes processes framework that deals with news arrival in FOREX market. This is the main original contribution of this thesis and all the material presented here is original.

In Section 6.1 we present an analysis of news impact on market activity, while in Section 6.2 we investigate the relation between news impact and how much news content matches market expectations.

Sections 6.3-6.4 introduce an extension of the Hawkes models presented so far, that aims at including in the description of market activity also news effects.

In Sections 6.5-6.6 we discuss the results of the application of the news model to real data and its capabilities of reproducing the observed behaviour of activity in correspondence of news arrivals.

Finally, in Section 6.6.3 we analyse the relationship between the estimated parameters of our model and the surprise in the news content.

6.1 News impact

Activity in all financial markets is sensitive to external events such as companies earnings publications and economic indicators updates. In particular, the FOREX market is very responsive to macroeconomic data announcements, much more than equity markets. This is not surprising, since the value of a currency is tightly linked to the economy of the country that backs it and, even more, to the monetary policy of central banks.

Before introducing our model it is important to analyse in some detail the impact of news on market activity. In this section we investigate this phenomenon in our FOREX database. In particular, our analysis develops along two complementary lines:

- Relation between bursts in activity and news release: what fraction of the observed bursts in market activity effectively corresponds to news releases?
- Impact of news on activity: what kind of increase in activity is observed after a news arrival?

Subsections 6.1.1 and 6.1.2 are devoted to the answers to these questions.

The news database was presented in Chapter 3. In our analysis we retained only the news with importance rated "Medium" and "High". We already noted that more than one news can be released at the same time. Usually they are different indicators of the same macroeconomic variable (e.g. employment). Since it is problematic to deal with simultaneous events, when there are multiple announcements we retained only the most important one. The ranking is performed first by using the "Importance" field in the database

and then according to the highest relative difference between Actual and Forecast values. Usually there is only an "High" importance news per group of simultaneous events, so this filtering should not alter too much the information on impact.

We restricted our analysis to London working hours (07:30-16:30 local time), and in the following we refer to this time window as a "day". One reason behind this choice is that, although some news are released during the London night (e.g. Bank of Japan interest rate announcements), market activity during the night is much lower and so modest changes in activity due to noise could be interpreted as burst in activity, resulting in flawed results. A second motivation is that for our modelling purposes we are interested mainly in London working hours and so it is desirable to have a picture of the impact of news in this window. After all these operations we are left with a total of 723 news.

We start our analysis by subdividing each day in intervals of the same length, say 1 min, and by associating to the centres t_i of these intervals the number of market events N_i registered in the corresponding interval. As in Chapter 5, we take as market event a change in any of the best quotes. Then, we calculate a Simple Moving Average (SMA_n) over the previous n intervals of the number of events. The $\mathrm{SMA}_n(t_i)$ is our measure of the expected activity for the interval t_i. Note that the events associated to t_i do not enter in the calculation of $\mathrm{SMA}_n(t_i)$. Finally, we define θ_i as the ratio between the observed number of events at t_i and the expected number of events, namely:

$$\theta_i = \frac{N_i}{\mathrm{SMA}_n(t_i)} \tag{6.1}$$

6.1.1 Relation between bursts in activity and news release

To answer the first question, namely what is the relation between observed spikes in market activity and news releases, we fixed four threshold values θ_t of θ, we calculated the number of θ_i above threshold, and then we counted the number of "anomalous" events that effectively corresponds to the release of a news. In Tables 6.1 and 6.2 the obtained results are summarized. We performed the analysis both with a 1 min interval and a moving average over a period of 20 min and with a 5 min bin and moving average over 30 min. In the first case a news is considered related to the anomalous event if it was released not more than 5 min before, for the second case this tolerance was set to 10 min.

Number of anomalous events per currency pair

		$\theta_t = 2$	$\theta_t = 3$	$\theta_t = 5$	$\theta_t = 7$
EUR/USD	*1 min*	4853	744	79	15
	5 min	585	75	7	2
EUR/JPY	*1 min*	6354	1030	119	28
	5 min	783	105	14	4
USD/JPY	*1 min*	11626	3050	444	131
	5 min	1376	280	47	16

Table 6.1: Number of anomalous events for different values of the threshold and for the three currency pairs.

Looking at Tables 6.1 and 6.2, we first note that the choice of the interval length has a sizeable impact on the results: a smaller bin size detects many more events, even considering that there are five times more one minute intervals than five minute intervals, most of which are not related to news. On the other hand, a larger interval length captures less events, but a higher percentage of them could be linked to news. Given the very quick reaction of modern electronic markets to market events, it seems that setting a small interval size, e.g. 1 min, it's more meaningful. In fact, it is common to have spikes of activity of very short duration and using a too large bin size in these cases averages out the bursts and results in a fail to detect it. Moreover, external news are not the only cause for these activity bursts: they can be

Percentage of anomalies related to news

		$\theta_t = 2$	$\theta_t = 3$	$\theta_t = 5$	$\theta_t = 7$
EUR/USD	*1 min*	9.1	17.1	34.2	46.7
	5 min	21.4	44.0	71.4	100.0
EUR/JPY	*1 min*	8.6	17.1	29.4	25.0
	5 min	20.0	42.8	71.4	75.0
USD/JPY	*1 min*	6.5	10.5	22.1	35.1
	5 min	17.0	33.6	76.6	81.2

Table 6.2: Percentage of anomalous activity attributable to news arrivals for different thresholds.

also generated endogenously, for example as reactions to bigger than usual movement in the price, and they can also be triggered by events half-way between news and fully endogenous events, signalled by so-called technical indicators (resistance, support, crossing of moving averages, etc.). Especially in the FOREX market, technical analysis is widely employed by traders.

To conclude this first part of the analysis, we considered also the effect of news anticipation. Since all macroeconomic news are released at scheduled times known to all market participants, it is possible to observe an increase in activity before the actual release. This could sometimes be the only effect of the news if the published value matches the forecast, i.e there are no surprises. Table 6.3 reports the percentage of activity spikes, as previously defined, related to news including the possibility of news anticipation. The difference here is that the tolerance interval extends both in the past and in the future with respect to the burst time. The tolerance was set to 5 min in both directions (10 min total) for the 1 min bin and to 10 min in both direction for the 5 min bin.

Finally, Table 6.4 shows the percentage of events explained by forthcoming news alone. It seems that there is a small part of the activity spikes that is in fact related to news anticipation. Overall, it emerges

Percentage of anomalies related to news including anticipation

		$\theta_t = 2$	$\theta_t = 3$	$\theta_t = 5$	$\theta_t = 7$
EUR/USD	*1 min*	11.7	20.4	35.4	53.3
	5 min	24.6	46.7	71.4	100.0
EUR/JPY	*1 min*	11.1	19.4	30.2	25.0
	5 min	24.4	42.8	71.4	75.0
USD/JPY	*1 min*	9.9	14.2	26.1	37.4
	5 min	23.8	40.7	78.7	81.2

Table 6.3: Percentage of anomalous activity attributable to news, including the effect of a forthcoming news, arrivals for different thresholds.

that on the FOREX market news play an important role, a non negligible fraction of the observed bursts in activity are related to news. This is an aspect to take into account in building a model of market activity.

6.1.2 Impact of news on activity

Concerning the second question, i.e. what is the reaction of market activity after a news release, in Figure 6.1 the histogram of the ratios θ that follow a news is reported for the three currency pairs. The ratio θ calculated after the release of a news is our measure of the news impact on market activity. We observe that there is always a number of θ that are in fact less than one, meaning that the news had no effect on activity or even a dumping effect. This is probably related to the fact that if the announcement of the

Percentage of anomalies related only to news anticipation

		$\theta_t = 2$	$\theta_t = 3$	$\theta_t = 5$	$\theta_t = 7$
EUR/USD	*1 min*	2.6	3.6	3.8	6.7
	5 min	3.2	2.7	0.0	0.0
EUR/JPY	*1 min*	2.5	2.4	0.8	0.0
	5 min	4.3	0.8	0.4	0.0
USD/JPY	*1 min*	3.7	4.0	4.0	2.3
	5 min	7.4	7.8	2.1	0.0

Table 6.4: Percentage of anomalous activity attributable only to news anticipation for different thresholds.

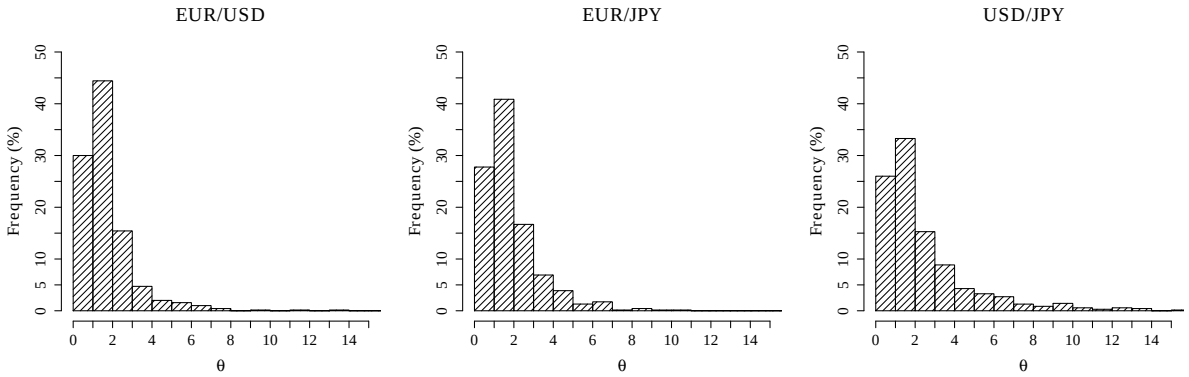

Figure 6.1: Histograms of the ratios θ that follow a news for all the three currency pairs.

latest figures for a certain economic indicator meets expectations, i.e. there is no surprise on the market, then there is little revision of market participants' strategies and hence no increase in activity, or in fact a decrease due to high activity concentrating before the announcement. We analysed the effect of all the news of the sample but also of specific subcategories: High importance only, EUR, USD or JPY related only. Figures 6.2 and 6.3 show the complements of the empirical cumulative distribution of the ratios θ that follow a news arrival for different categories of news. For comparison, we also report for each pair the distribution of $20,000$ θ chosen at random times, hence not necessarily following a news.

As expected, high importance news have actually a more pronounced effect on activity, signalled by the higher probability of high values of θ. We note also that the effect on USD/JPY is always somewhat higher than on the other pairs. This might be linked to USD/JPY being the less liquid of the three. Filtering the news by the currency field highlights the central role of the US economy: news marked as USD related have an higher impact than EUR related, regardless of the currency pair considered. We note also that USD news impact EUR/JPY as well, and also EUR news impact USD/JPY. Though expected, it is not automatic that a news related to EUR or USD should impact in the same way (or even more) USD/JPY or EUR/JPY respectively as they do on EUR/USD. This is due to EUR and USD being the two most important currencies, and the respective central banks the most influential ones. The greater impact on EUR/JPY and USD/JPY as compared to EUR/USD is probably an effect of the lower liquidity of the firsts. Finally, it should be stressed that most important JPY news are released outside the London working hours, therefore the JPY curves in Figure 6.3 are calculated on a small sample which is not fully representative of the JPY news population.

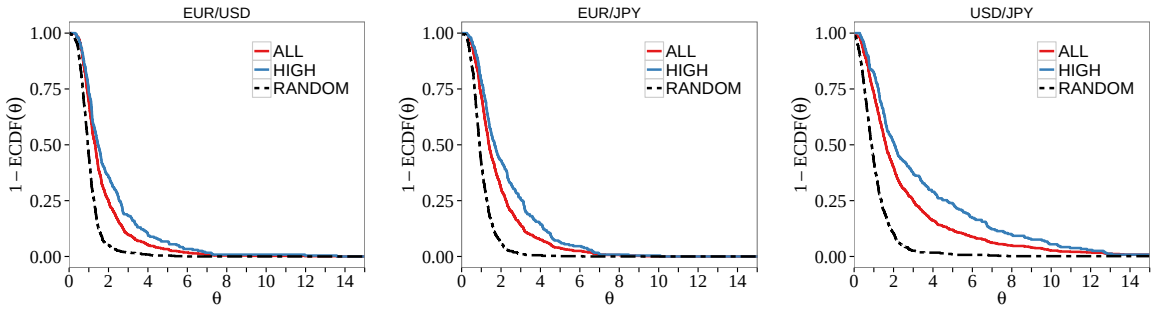

Figure 6.2: Complement of the Empirical Cumulative Distribution of θ after a news for the three currency pairs. "ALL" indicates the distribution of θ after all the news we considered (High and Medium importance), "HIGH" denotes the distribution of θ restricted to High importance news only. The dashed line "RANDOM" is the distribution of a sample of random θ, not necessarily following a news.

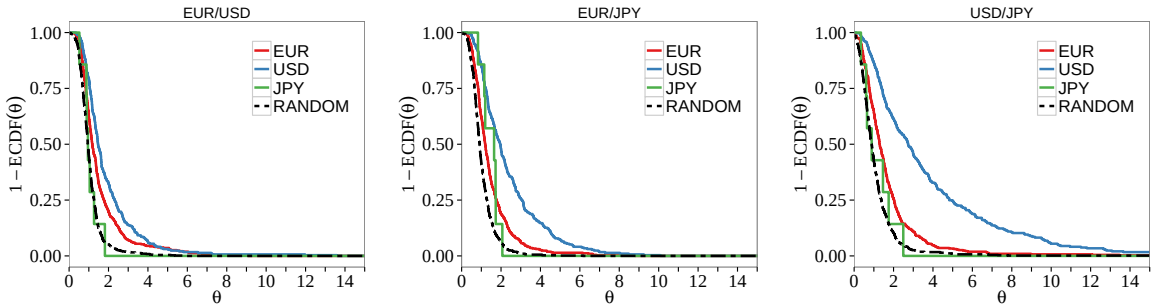

Figure 6.3: Complement of the Empirical Cumulative Distribution of θ after a news for the three currency pairs. "EUR" indicates the distribution of θ after news labelled "EUR" in the database, i.e. whose content is mainly related to Eurozone economy. Analogously, "USD" denotes the distribution of θ after news marked "USD", and "JPY" designates the distribution of θ after news marked "JPY". The dashed line "RANDOM" is the distribution of a sample of random θ, not necessarily following a news. For the "JPY" line is important to note that we are considering only London working hours and most of JPY-related news fall outside this interval.

6.2 Surprise

So far, we have analysed the impact of news disregarding their actual content, that is the value of the economic variable to be released. It is intuitive that, generally speaking, the impact of a news is related both to the overall importance assigned to the indicator and to the difference between the value the market expect and the one that is published. The expectation of the market for a certain indicator is summarized by the so called "Analysts consensus", reported in our news dataset in the "Forecast" field. This is an average of analysts' forecasts published by mayor news company such as Reuters, obtained by polling a number of market analysts.

In this section we try to investigate and formalize the impact of the difference between expectations and realizations, which we call "surprise". First we must note that not every news concerns the release of a quantitative variable. Press conferences of central banks governors are an example: although their effect could be very high, no measurable data is given. In this cases it is difficult, if not impossible, to assign a numeric value to the surprise that came with the news. Hence, for this analysis, we removed from the database all the lines that do not contain a numeric value for an economic indicator. Moreover, we removed also those lines for which a "Forecast" value was not given. In the end, starting with 723 news, we are left with a total of 559 news or about 77% of the original sample.

After this data cleaning, for each news that remains we have the Actual and the Forecast value of the indicator I, which we denote I_A and I_F respectively. The measures of surprise that we use are the absolute surprise

$$S_{\text{abs}} = |I_A - I_F|, \tag{6.2}$$

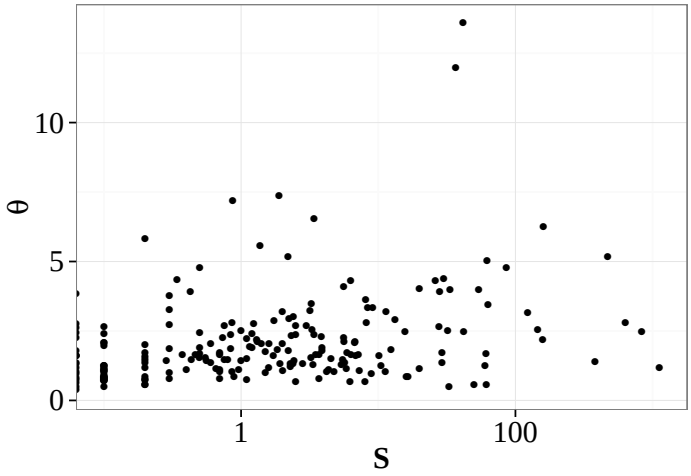

Figure 6.4: Scatterplot of θ against our measure of news surprise S. The figure refers to EUR/USD data for θ.

and the relative surprise

$$S_{\text{rel}} = |\frac{I_A - I_F}{I_F}|. \tag{6.3}$$

In principle one can distinguish between "good news" and "bad news", i.e. if the indicator was better or worse than expected. However this is difficult to do since the indicator are very diverse and whether a positive increment is good or bad depends on the type of indicator. Therefore we chose to use the absolute value of the difference. The use of one measure or the other is not immediate. On one hand, in fact, the indicators span a wide range, some are of the order of unity other are given in thousands or in percentage and hence the most reasonable choice seems the relative surprise. On the other hand, for some indicators, such as unemployment rate, that are given in percentage, even a small absolute change is weighted by the market.

For these reasons we chose the surprise indicator to be the absolute surprise for indicators given in percentage and the relative surprise in all other cases. Now we want to examine the relation between the surprise we have defined and impact, represented by θ.

In Figure 6.4 the scatterplot of θ vs. surprise is reported for EUR/USD. The plot is not very informative, the relation between S and θ appears very noisy. The plot looks almost the same across the different pairs and considering only high importance news helps little. This outcome is related to the over-simplistic measure we adopted, which is not able alone to highlight a clear dependence between the two variables. However, if we consider instead $\Pr(\theta > 1|S)$, we obtain a more significant result. In Figure 6.5 the probability of observing $\theta > 1$ after a news conditional on the value of S is reported for the three pairs. To obtain the graphs of figure 6.5 we proceeded as follows:

1. We calculated the empirical quintiles of the surprise distribution in our dataset, only news with $S \neq 0$ were considered.

2. We used these values to group the news according to their surprise. A sixth group was formed with the news with exactly zero surprise.

3. Then, $\Pr(\theta > 1|S)$ was estimated for each group as the ratio of news with $\theta > 1$ on the total number of news in the group

For each point in the graph the abscissa is the average surprise in each quintile. Now the relation between impact and surprise is apparent. For EUR/USD and EUR/JPY we obtain an almost monotone trend, while USD/JPY presents a more noisy pattern. It seems that even this very simple measure of surprise, that does not take into account the specificity of each indicator, is nevertheless able to capture some relevant information.

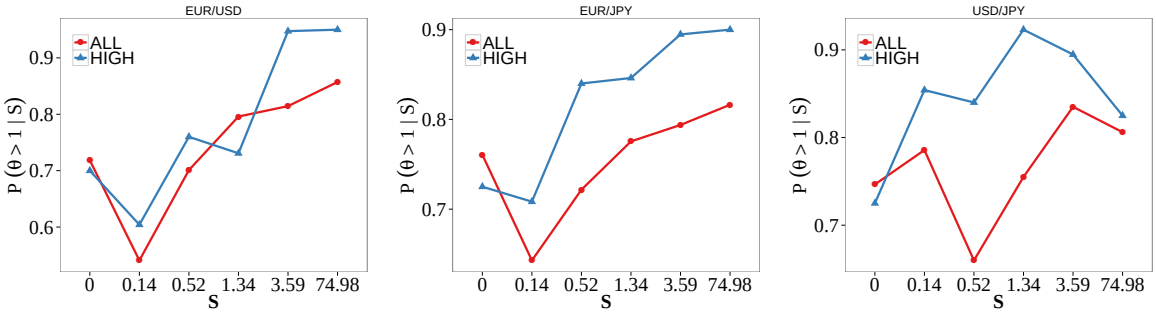

Figure 6.5: Conditional probability of observing $\theta > 1$ given the news surprise. Values are shown for all news we considered and for High importance news only.

6.3 The model

We now present an extension of the Hawkes models presented so far that includes a news-related (exogenous) term. The model aims at reproducing in a Hawkes framework the impact of news arrival on market activity. It is important to note that the news process is considered as given, that is, the model is not meant to describe the news process itself. The news process is completely deterministic in this context. As we said, this is actually also the real case in FX market, where the news are announced at predetermined times.

Let $\{n_j\}_{j\in\mathbb{N}}$ indicate the news process, n_j is thus the time at which j-th news is released. The intensity of the market activity process $\{t_i\}_{i\in\mathbb{N}}$ is then written

$$\begin{aligned}\lambda(t|\mathcal{F}_t) &= \mu + \int w(t-s)\mathrm{d}N_s + \int w_N(t-n)\mathrm{d}N_n^{\text{news}} \\ &= \mu + \sum_{t_i<t} w(t-t_i) + \sum_{n_j<t} w_N(t-n_j)\end{aligned} \tag{6.4}$$

where $w(t)$ is the endogenous self-exciting kernel and $w_N(t)$ is the exogenous kernel that accounts for the news-induced excitement. In essence, we are considering only one component of a bivariate process.

We now derive the condition for stationarity. As in equation (2.20), we take the unconditional expectation on both side of (6.4)

$$\begin{aligned}\overline{\lambda} &= \mu + \int_{-\infty}^{t} w(t-s)\mathbb{E}\left[\mathrm{d}N(s)\right] + \int_{-\infty}^{t} w_N(t-n)\mathbb{E}\left[\mathrm{d}N^{\text{news}}(n)\right] = \\ &= \mu + \int_{-\infty}^{t} w(t-s)\mathbb{E}\left[\lambda(s)\right]\mathrm{d}s + \int_{-\infty}^{t} w_N(t-n)\mathbb{E}\left[\lambda_N(n)\right]\mathrm{d}n = \\ &= \mu + \overline{\lambda}\int_{-\infty}^{t} w(t-s)\mathrm{d}s + \overline{\lambda}_N\int_{-\infty}^{t} w_N(t-n)\mathrm{d}n = \\ &= \mu + \overline{\lambda}\int_{0}^{\infty} w(\tau)\mathrm{d}\tau + \overline{\lambda}_N\int_{0}^{\infty} w_N(\tau)\mathrm{d}\tau\end{aligned} \tag{6.5}$$

solving for $\overline{\lambda}$, we finally get

$$\overline{\lambda} = \frac{\mu + \overline{\lambda}_N\int_0^\infty w_N(\tau)\mathrm{d}\tau}{1 - \int_0^\infty w(\tau)\mathrm{d}\tau}. \tag{6.6}$$

Provided that $\overline{\lambda}_N$ exists and is finite, the condition of stationarity is thus the same of the self-exciting only case, namely $\int_0^\infty w(\tau)\mathrm{d}\tau < 1$. In particular, it is not necessary that also $\int_0^\infty w_N(\tau)\mathrm{d}\tau < 1$ in order to have stationarity of market activity.

	EURUSD	EURJPY	USDJPY
Avg. number of best-quote-change events	17579	11199	7322

Table 6.5: Average number of events in each window of three hours centred on a news.

6.4 Kernel specification

To actually test model (6.4) it is necessary to specify the functional form of $w(t)$ and $w_N(t)$. As we have seen in the previous chapters, the double exponential kernel and the quasi-power law kernel give good results for the endogenous component, hence we adopted these form also here. For the exogenous kernel w_N we chose a single exponential specification of the type

$$w_N(t) = \alpha_N e^{-\beta_N t}. \tag{6.7}$$

This choice has the advantage that the two parameters have a clear interpretation, α_N gives the magnitude and β_N fixes the time scale. Moreover, it seems reasonable, at least as a first approximation, that the news impact could be described by a single timescale. Besides, as already pointed out in Chapter 2, exponential specifications have nice computational properties.

We thus test two forms of the intensity, which differs only for the form of the endogenous kernel $w(t)$, namely

$$\lambda_{\text{DE}}(t) = \mu + \int \alpha_A e^{-\beta_N(t-s)} + \alpha_B e^{-\beta_B(t-s)} \mathrm{d}N_s + \int \alpha_N e^{-\beta_N(t-u)} \mathrm{d}N_u^{\text{news}} \tag{6.8}$$

$$\lambda_{\text{PL}}(t) = \mu + \int \frac{n}{Z} \{ \sum_{k=0}^{M} (\tau_0 m^k)^{-p} e^{-\frac{t-s}{\tau_0 m^k}} - S e^{-\frac{t-s}{\frac{\tau_0}{m}}} \} \mathrm{d}N_s + \int \alpha_N e^{-\beta_N(t-u)} \mathrm{d}N_u^{\text{news}} \tag{6.9}$$

"DE" stands for Double Exponential and "PL" for power law as in the previous chapters. In the first case we have a total of 7 parameters and 6 in the second. The expression for the log-likelihood is derived using (2.32), and thanks to the exponential kernels can be computed recursively.

6.5 Estmation and empirical test of the model

We estimated and tested our model on the series of best quote changes for the three currency pairs in our database. The news are those described in section 6.1 of this chapter. On each day there are on average three news announcements. Since the news are very diverse from each other and instead our model weights all the news equally, we decided to calibrate the model on a window were only one news is present. The procedure we followed is summarized below.

1. From our news database we individuate those news events that are "isolated". That is no other news in the database happens 1.5 hours before and 1.5 hours after. This 3 hour window allows to have a reasonable number of best-quote-change events for likelihood estimation.
2. For each pair of currencies we select the best-quote-change events that fall in the time windows created at point 1.
3. Parameter estimation for each version of the model is performed in each window.

We note that in this way the news event happens always at $t = 5400$ s, i.e. at half interval. Proceeding in this way, we obtained a total of 267 isolated news. The news are the same for all the pairs. The average number of events in each window for each pair is reported in Table 6.5.

For the estimation of the parameters we used the same approach of the previous chapters. To find the values of the parameters that maximize $\ln \mathcal{L}$ we employed R *nlminb* function. Ten different combination of starting values were tried and the best result was retained.

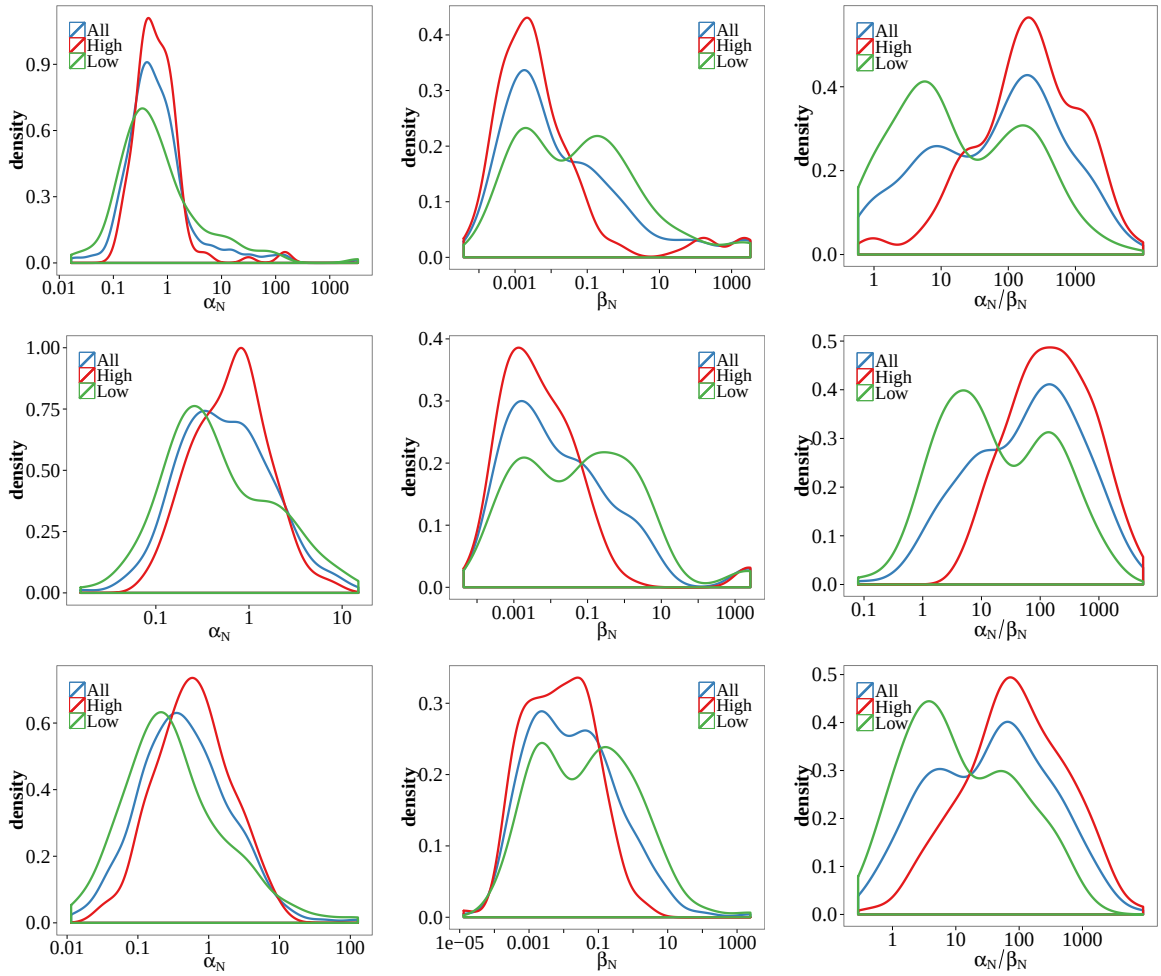

Figure 6.6: Estimated probability density of values of α_N, β_N and α_N/β_N. Each row of graphs corresponds to a different pair. From top to bottom we have: EUR/USD, EUR/JPY and USD/JPY. The curves labelled "High" and "Low" are the probability densities corresponding to high impact news ($\theta > \theta_{\text{median}}$) and low impact ($\theta \leq \theta_{\text{median}}$) news respectively. "All" instead comprises all values of the impact. The scale of the x axis is logarithmic, hence values of α_N and of the ratio α_N/β_N exactly equal to zero are not plotted. Their number is reported in Table 6.6.

6.6 Results

To analyse the performance of our model we divided the news on which it was tested in two groups: high impact news and low impact news. The groups are made of the news to which corresponds a θ greater than the sample median of θ and, respectively, lower. In Figure 6.6 we present the distribution of the parameters α_N and β_N we found with the model that uses the power law endogenous kernel. We will consider mainly this model, which as we shall see, gives better results. Moreover, the distribution of the news parameters is very similar between the two models.

First, we note that the news-related parameters have a very broad distribution. This is due both to the very different effects that news produce and to the uncertainty in the estimation. As a matter of fact, uncertainty on the news-related parameters, calculated from the inverse of the Hessian matrix, is very high, with averages of about 50% for both parameters. Despite the high uncertainty, however, the distributions in the case of high impact news and low impact news are clearly distinguishable one from each other. This is especially true for the ratio α_N/β_N, where two peaks emerges one between 100 and 1000 for high impact news an the other between 1 and 10 for low impact news. Another small peak is present at zero (not shown in the figures). In fact some estimates of α_N are exactly zero. Their number along with the corresponding average impact is detailed in Table 6.6.

	Number of news with $\alpha_N = 0$	$\overline{\theta}$
EUR/USD	20	1.15
EUR/JPY	22	1.06
USD/JPY	16	0.54

Table 6.6: Number of α_N found exactly equal to zero for each pair. The average value of the corresponding θ, $\overline{\theta}$, is also reported.

	parameter	t	df	$p-$value
EUR/USD	α_N	1.03	134.10	0.31
	β_N	0.34	258.95	0.73
	$\frac{\alpha_N}{\beta_N}$	-3.98	183.47	$1.0{\cdot}10^{-4}$
EUR/JPY	α_N	0.47	206.51	0.64
	β_N	0.25	263.72	0.80
	$\frac{\alpha_N}{\beta_N}$	-4.82	152.11	$3.4{\cdot}10^{-6}$
USD/JPY	α_N	0.95	138.02	0.34
	β_N	1.11	133.00	0.27
	$\frac{\alpha_N}{\beta_N}$	-3.72	138.98	$2.8{\cdot}10^{-4}$

Table 6.7: Results of the t-test of the hypothesis that the parameters for high impact and low impact news have the same mean. The test is two sided and the Welch approximation is used for the degrees of freedom (df) to account for unequal variances in the two samples. We report the value of the t-statistic along with the computed $p-$value. The test was performed using R built-in function `t.test`.

The considerations we made on Figure 6.6 are supported also by the results of the t-test of the hypothesis that the means for each parameter are the same for high and low impact news, whose results are summarized in Table 6.7. The ratio α_N/β_N emerges as the variable that better captures the news characteristics, with very small p-values favouring the rejection of the hypothesis that the two groups share the same mean. Its value is in fact linked to the news impact much more clearly than that of the single parameter α_N and β_N.

When a news produces little or no impact, from the model point of view this can be interpreted with a small value of the amplitude α_N, but also with a large value of β_N. The latter imposes a very fast decay of the exogenous intensity component and hence produces an almost identical result to small values of the amplitude. Actually, the characteristic times β_N^{-1} associated with high impact news have typical values of several minutes, while those relative to low impact news of only 0.1s or less. The ratio α_N/β_N incorporates both these aspects.

We said that the news parameter estimates are subject to high uncertainty. Then, to check that nevertheless we obtained meaningful results, we estimate the same parameters on a set of simulated paths with no news. In particular, we generated by simulation several paths of a Hawkes process with intensity

$$\lambda(t) = \mu + \int w(t-s)\mathrm{d}N_s = \mu + \sum_{t_i<t} w(t-t_i) \tag{6.10}$$

using for $w(t)$ both the double exponential and the power law kernel. Five different sets of parameters, reported in Table 6.8, were used for each kernel, and with each set 140 simulated paths were produced, for a total of 700 simulation per kernel. The parameters for the simulation are chosen to be very close to the estimated values obtained for the endogenous parameters with maximum likelihood estimation, so that the simulated paths have the same characteristics of the observed endogenous process. The time length of the simulated interval was set equal to the three hour window we used with real data.

	Power law kernel				Double Exponential kernel				
	μ	n	τ_0	p	μ	α_A	β_A	α_B	β_B
Set 1	0.25	0.85	0.22	1.65	0.31	1.20	2.30	0.02	0.05
Set 2	0.07	0.91	0.17	1.41	0.13	1.40	2.80	0.03	0.08
Set 3	0.01	1.02	0.28	1.23	0.18	0.90	2.00	0.04	0.09
Set 4	0.31	0.73	0.25	1.72	0.07	1.07	1.91	0.02	0.07
Set 5	0.11	0.87	0.24	1.88	0.12	1.84	3.33	0.03	0.13

Table 6.8: Parameters sets used for the simulations. Every set was used to simulate 140 paths. A total of 700 paths per kernel was produced.

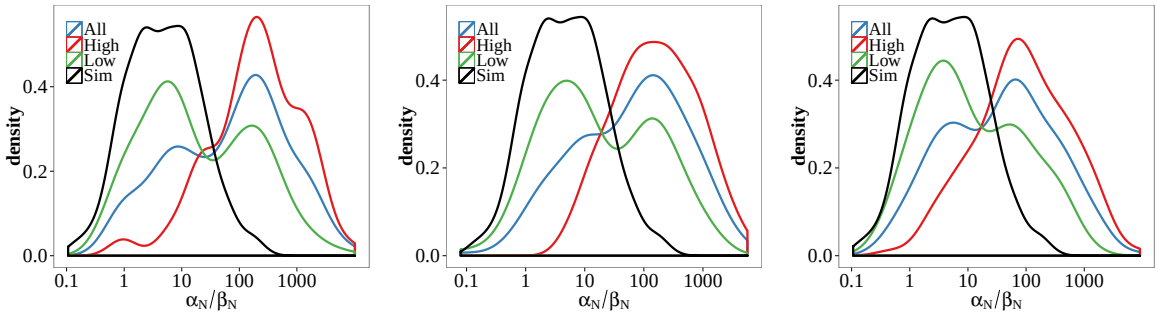

Figure 6.7: Comparison of estimated probability density of the ratio α_N/β_N between values estimated from real data and those estimated from simulations where no news is present. Results are those obtained using the power law endogenous kernel. From left to right: EUR/USD, EUR/JPY and USD/JPY.

We then estimated the best parameters of model of equations (6.9) and (6.8) on these simulated data in order to study the behaviour of the model in absence of news. Figure 6.7 shows the distribution of the ratio α_N/β_N we get from simulated data compared to the ones obtained with real data. Values exactly equal to zero are not shown and are instead reported in Table 6.9.

We note that the peak in the distribution of the ratio α_N/β_N for the simulated sample is located in the same region as the one of low impact news, while that for high impact news is clearly separated. Also the fraction of ratios equal to zero is much higher for parameters estimated on simulations. We therefore conclude that the values given by the model for news with significant impact emerge distinctly from noise.

Fraction of ratios equal to zero (%)			
	High	Low	Tot
EUR/USD	2.25	5.24	7.49
EUR/JPY	1.50	7.11	8.61
USD/JPY	0.00	5.99	5.99
SIM	-	-	23.28

Table 6.9: Fraction of the ratios α_N/β_N exactly equal to zero, for the three currency pairs and for the estimation on simulated data. The fraction is calculated over the whole sample, i.e. 267 values for each pair and 700 for the simulated data. "High" indicates the fraction of the total number of ratios which are equal to zero and correspond to high impact news. "Low" is he same figure for low impact news.

6.6.1 Simulations of the process with the extended model

To get a visual impression of the performance of the model, we simulated a process with the estimated parameters on the same time window and we compare it to the actual observation. In Figure 6.8 we compare simulation from all the models with actual data in the case of a very relevant news, namely, the release of US change in non-farm payrolls which turned out to be much lower than expected. Clearly the models of chapter 5, that do not feature an exogenous term, fail to reproduce the trend of activity after an important news. We also note that the model (6.9), with the endogenous power law kernel, seems to reproduce pretty well both the magnitude and the temporal decay of news impact, while the one with the double exponential kernel overestimates the impact immediately after the news arrival. Thus, it appears that the choice of a single exponential kernel for the news component is adequate and that a good description of the endogenous process is needed in order to obtain a faithful reproduction of the news effect. Figure 6.9 is analogous, a high impact news on EUR/JPY is depicted. In this case the announcement concerned the update of the German ZEW survey of economic sentiment that was better than expected. Here we plot only the simulations with the power law kernel, with and without the news term. Figure 6.10 shows an high impact news on USD/JPY, here we focus on the $800\,s$ centred on the announcement. Figure 6.11 refers instead to the effects of the announcement of the Euro-Zone industrial new orders figures on USD/JPY. The figures were in line with expectations and had a scarce impact. Again, the model behaves nicely, in that it correctly captures the absence of impact in this case. It is interesting to note that the values of the amplitude obtained for the low-impact case are actually bigger than those for the high-impact cases. The fact that the model correctly reproduce the absence of impact is ascribable to the very low estimated characteristic time. To complete the analysis, in Figure 6.12 we

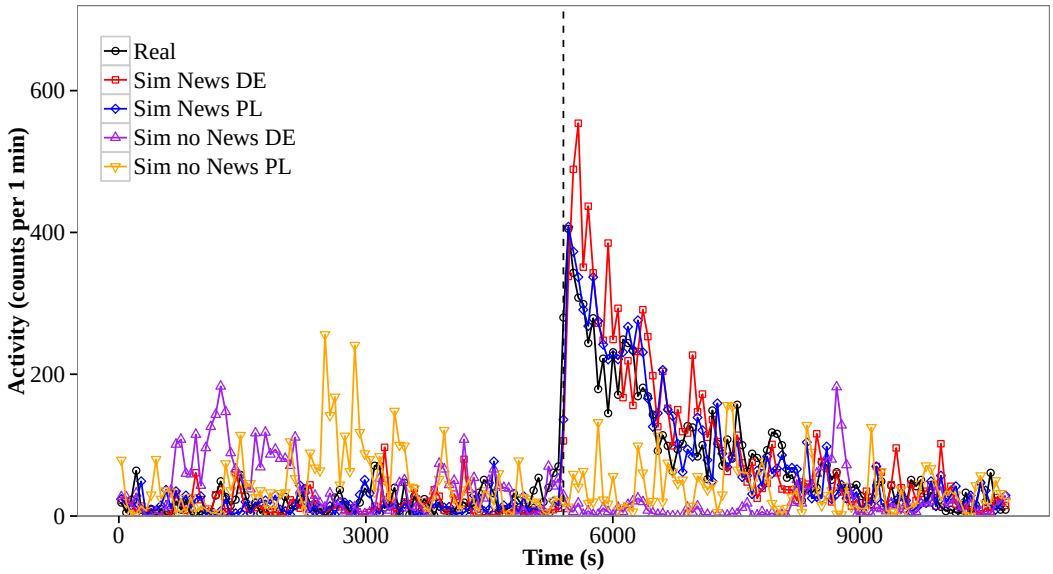

Figure 6.8: Actual and simulated activity measured as events per 1 min for EUR/USD on Friday April 06, when an important figure on US change in non-farm payrolls was released at 13:30. The figure produced an impact $\theta \approx 13$ and was much worse than expected (Surprise =44). The dashed line corresponds to the time of the announcement. Results from simulation of the models without news are also shown for comparison. Parameter estimates were $\alpha_N^{\mathbf{DE}} = 3.2$, $\beta_N^{\mathbf{DE}} = 1.6 \cdot 10^{-2}$, $\alpha_N^{\mathbf{PL}} = 1.45$, $\beta_N^{\mathbf{PL}} = 8.2 \cdot 10^{-4}$ (all values in s^{-1}).

consider a case where the impact was intermediate. Here the model captures only partially the effect of the news. In particular the effects immediately after the news is a bit underestimated.

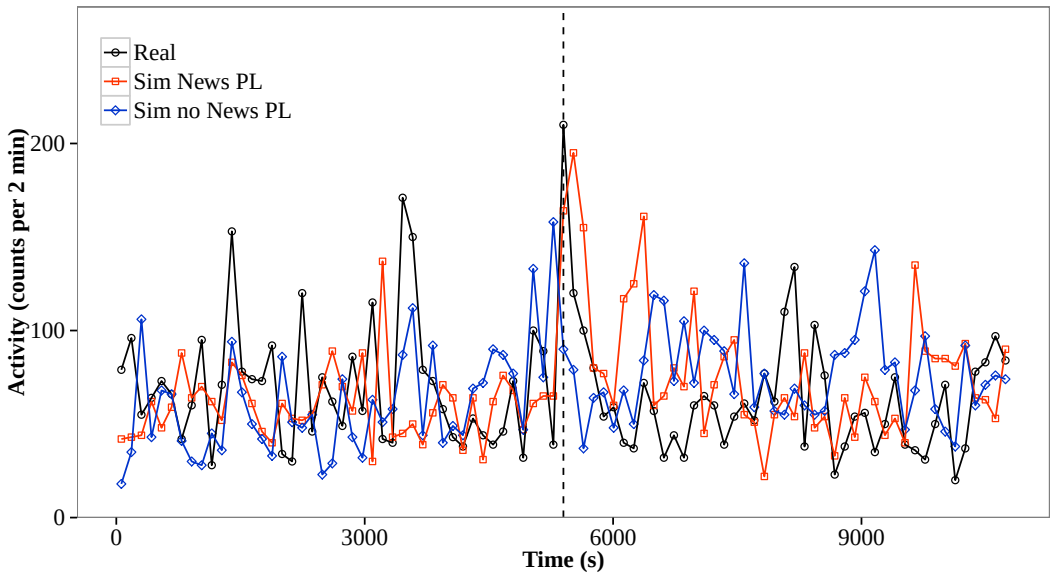

Figure 6.9: Actual and simulated activity measured as events per 2 min for EUR/JPY on Tuesday December 11 in correspondence of the announcement of the update of the German ZEW Survey of economic sentiment at 10:00. The figure produced an impact $\theta \approx 7$ and was much better than expected (Surprise =160). The dashed line corresponds to the time of the announcement. Parameter estimates were $\alpha_N^{\mathbf{DE}} = 0.80$, $\beta_N^{\mathbf{DE}} = 7.8 \cdot 10^{-3}$, $\alpha_N^{\mathbf{PL}} = 0.89$,$\beta_N^{\mathbf{PL}} = 9.0 \cdot 10^{-3}$ (all values in s^{-1}).

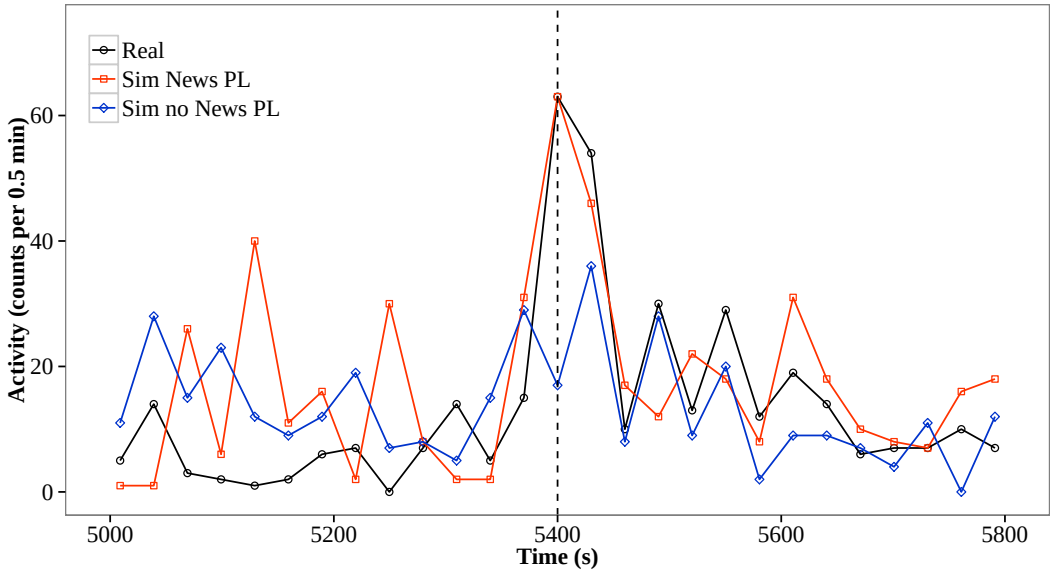

Figure 6.10: Actual and simulated activity measured as events per 30 s for USD/JPY on Wednesday December 5 in correspondence of the announcement of the update of the US ISM Non-Manufacutring Composite index at 15:00. The figure produced an impact $\theta \approx 9$ and was better than expected (Surprise =2.2). The dashed line corresponds to the time of the announcement. Only the $800s$ centred on the news are shown. Parameter estimates were $\alpha_N^{\mathbf{DE}} = 3.9$, $\beta_N^{\mathbf{DE}} = 1.1 \cdot 10^{-1}$, $\alpha_N^{\mathbf{PL}} = 3.16$,$\beta_N^{\mathbf{PL}} = 8.6 \cdot 10^{-2}$ (all values in s^{-1}).

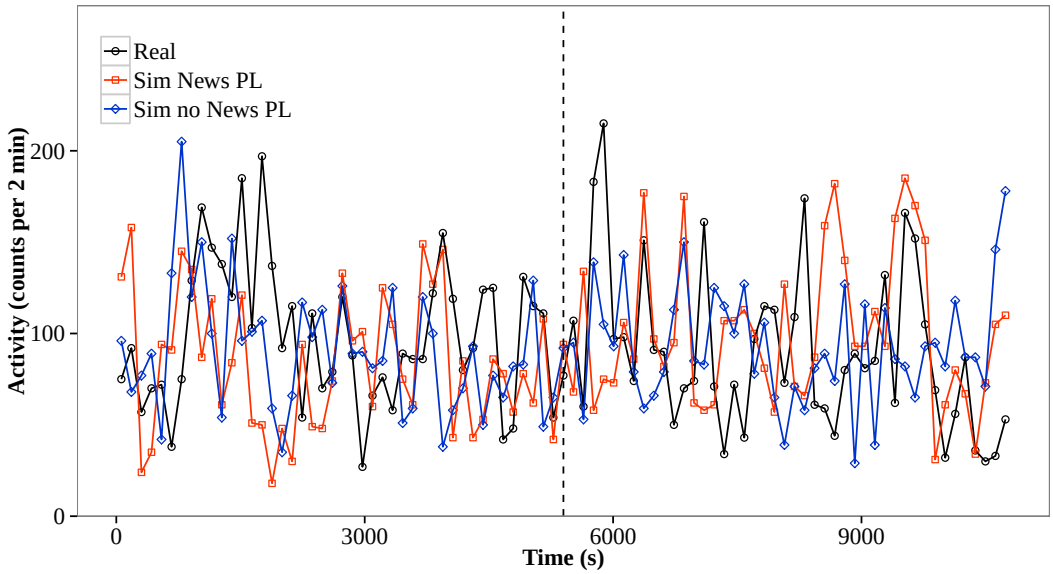

Figure 6.11: Actual and simulated activity measured as events per 2 min for USD/JPY on Thursday January 5 around the publication of Euro-Zone Industrial New Orders data. The value matched expectations and produced no increase in activity ($\theta \approx 0.8$). Moreover it mostly concerned EUR. The dashed line corresponds to the time of the announcement. Parameter estimates were $\alpha_N^{\mathbf{DE}} = 3.0$, $\beta_N^{\mathbf{DE}} = 7.0$, $\alpha_N^{\mathbf{PL}} = 2.72$, $\beta_N^{\mathbf{PL}} = 7.0$ (all values in s^{-1}).

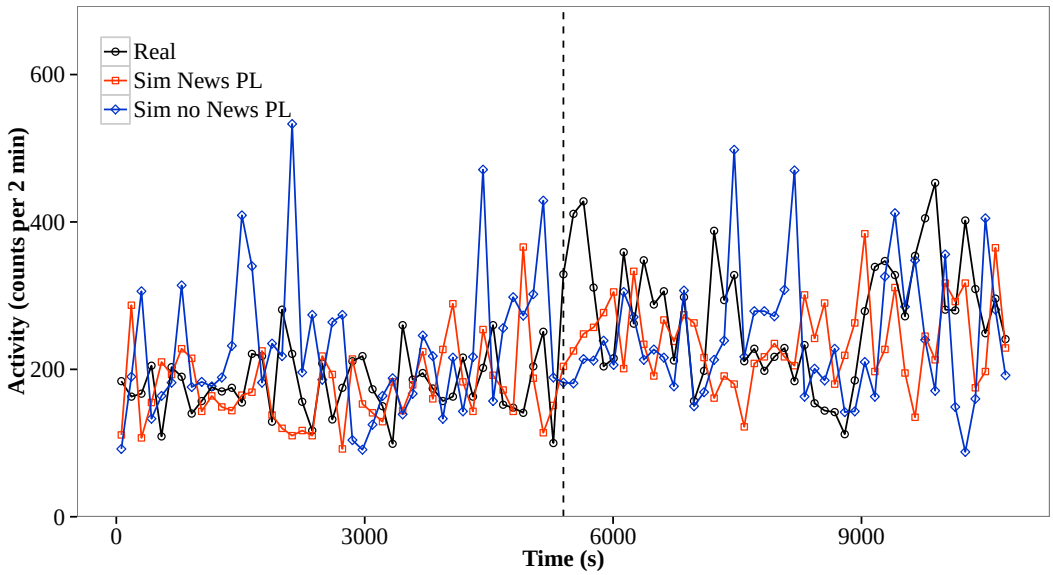

Figure 6.12: Actual and simulated activity measured as events per 2 min for EUR/USD on Tuesday February 28, the window is centred on the publication of US Durable Goods Orders. The news lead to a modest increase in activity ($\theta \approx 2.7$). The dashed line corresponds to the time of the announcement. Parameter estimates were $\alpha_N^{\mathbf{DE}} = 0.16$, $\beta_N^{\mathbf{DE}} = 4 \cdot 10^{-5}$, $\alpha_N^{\mathbf{PL}} = 0.16$, $\beta_N^{\mathbf{PL}} = 6 \cdot 10^{-5}$ (all values in s^{-1}).

6.6.2 Analysis of the endogenous and exogenous contributions to the intensity

In the model we presented, the total intensity $\lambda(t)$ is the sum of three contributions: the baseline intensity μ, the endogenous component $\int w(t-s)\mathrm{d}N_s$ and the exogenous component $\int w_N(t-u)\mathrm{d}N_u^{\text{news}}$. It is worth analysing how the contribution of each term varies over time. Figures 6.13 and 6.14 show the fraction of the intensity coming from each component as a function of time for two highly important news. The curves are computed calculating the components of $\lambda(t)$ from (6.9) with the estimated parameters. Values are computed every $0.1s$ and we used a simple moving average over 1000 points ($100s$) to smooth the lines.

After the news release, the contribution of the exogenous term rises quickly, while the weight of the endogenous one decreases. Also the contribution of the baseline intensity becomes negligible immediately after the announcement. The exogenous contribution then slowly decays towards zero and the endogenous and the baseline components regain their pre-news level. It is interesting to contrast these figures with Figure 6.15, where the contributions to $\lambda(t)$ are shown for the model without the news term for the same data of Figure 6.14. The increment of activity after the news is now attributed to the endogenous component. The contribution of the baseline intensity has a similar trend in the two cases, though we note that its contribution before the news is higher in the model with the exogenous term. We will discuss further this point in Section 6.7.

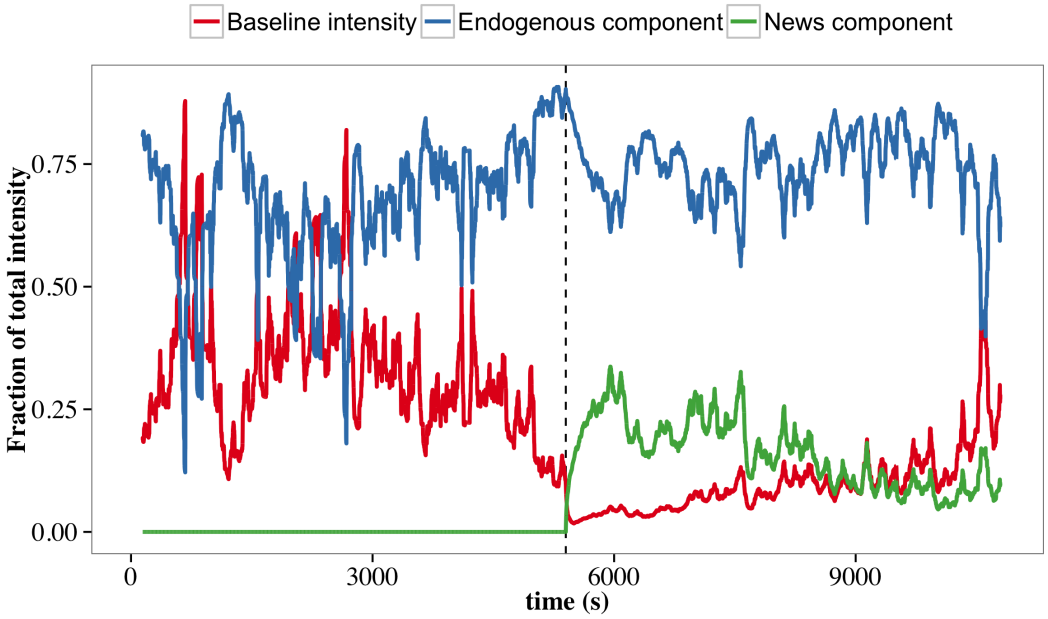

Figure 6.13: Components of the intensity $\lambda(t)$ as a function of time for an high impact news. The components sum to one and the curves are smoothed using a simple moving average over 100 s. The figure refers to USD/JPY, in the window centred on the announcement of the ISM manufacturing index (Tuesday 01 May 2012 at 14:00). The released value was better than expected ($S = 3.4$, $\theta = 11.8$).

In Figure 6.16 the same plots are proposed for two low impact news. It is possible to appreciate how, in this case, the exogenous component decays much faster and has a marginal role.

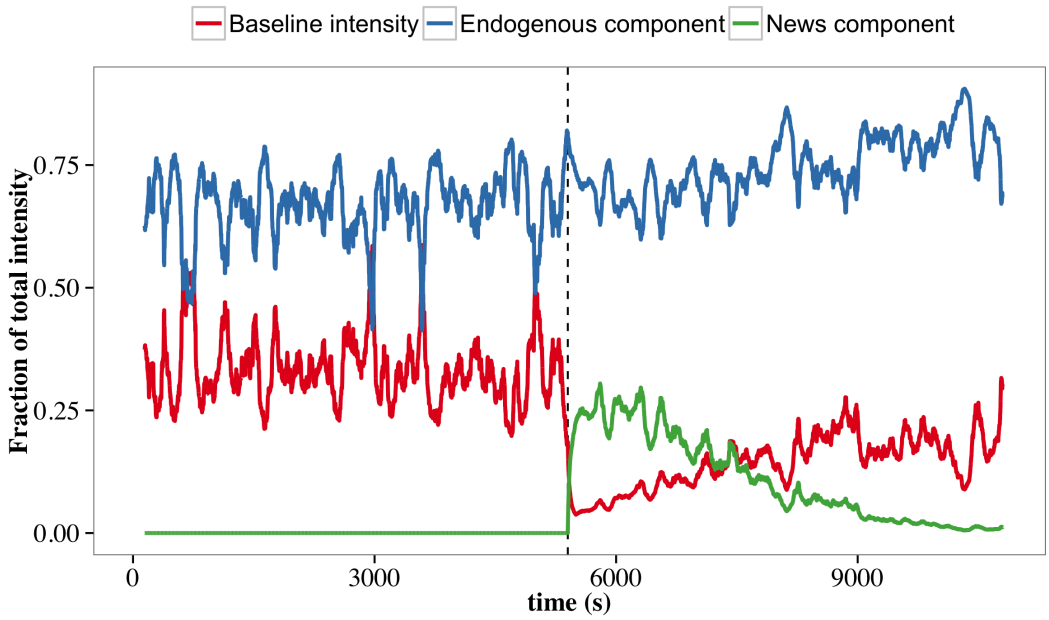

Figure 6.14: Components of the intensity $\lambda(t)$ as a function of time for an high impact news. The components sum to one and the curves are smoothed using a simple moving average over $100\,s$. The figure refers to EUR/JPY, in correspondence of the announcement of US new jobs data on Friday 02 November 2012 at 12:30. More jobs than expected were created ($S = 36$, $\theta = 10$).

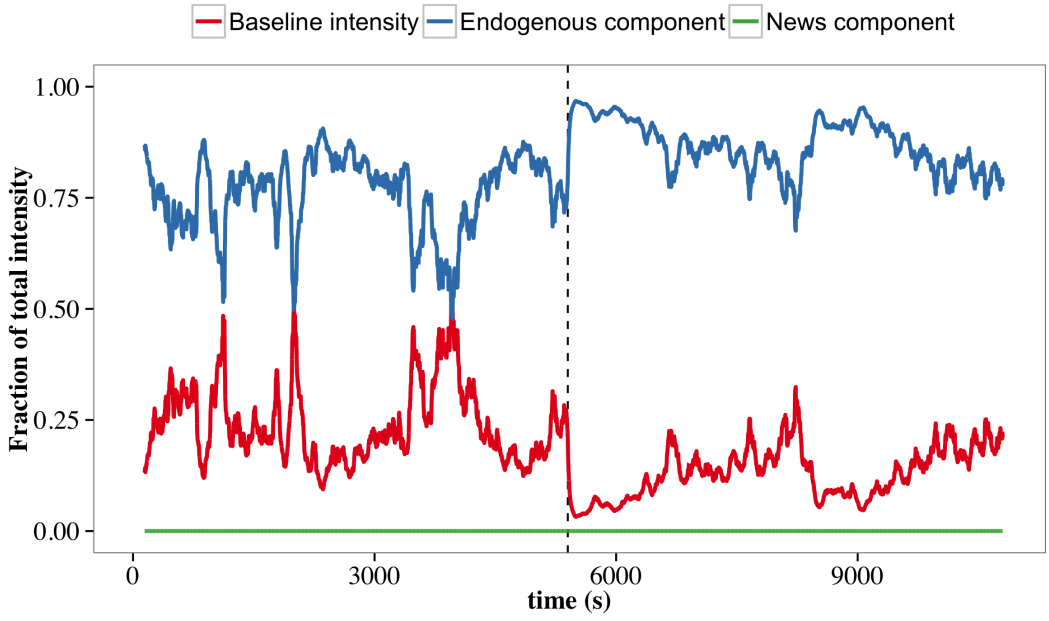

Figure 6.15: Components of the intensity $\lambda(t)$ as a function of time for the same news of Figure 6.14 but for the model without news term. The components sum to one and the curves are smoothed using a simple moving average over $100\,s$.

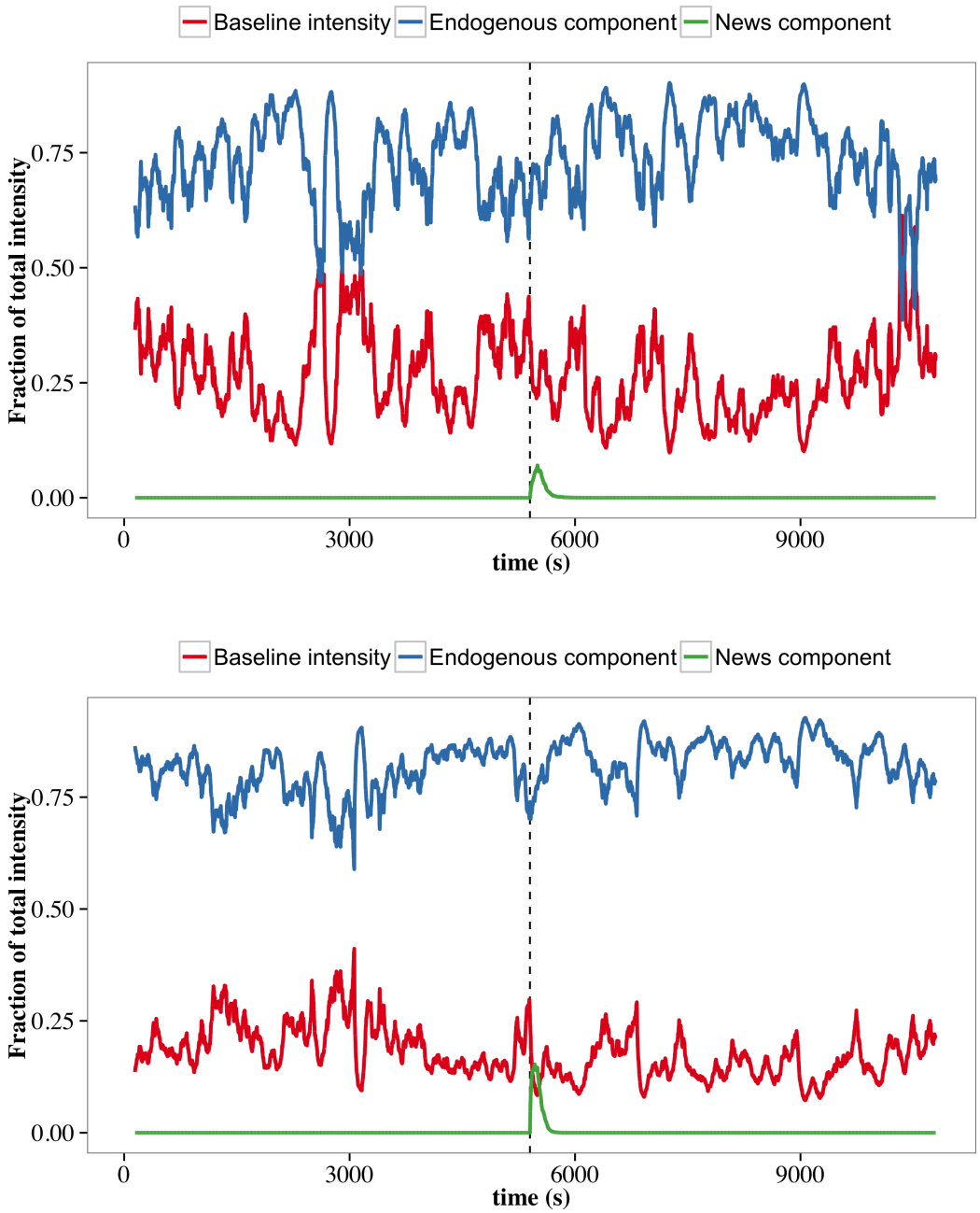

Figure 6.16: Components of the intensity $\lambda(t)$ as a function of time for two low impact announcements. The components sum to one and the curves are smoothed using a simple moving average over 100 s. The figure on top refers to USD/JPY, in the window centred on the announcement of the change in US factory orders (Wednesday 04 January 2012 at 15:00). The released value was in line with expectations ($S = 0.2$, $\theta = 1.1$). The figure on bottom refers to EUR/USD, in correspondence of the announcement of US new home sales data on Wednesday 26 September 2012 at 14:00. The value substantially matched forecasts ($S = 1.8$, $\theta = 1.8$).

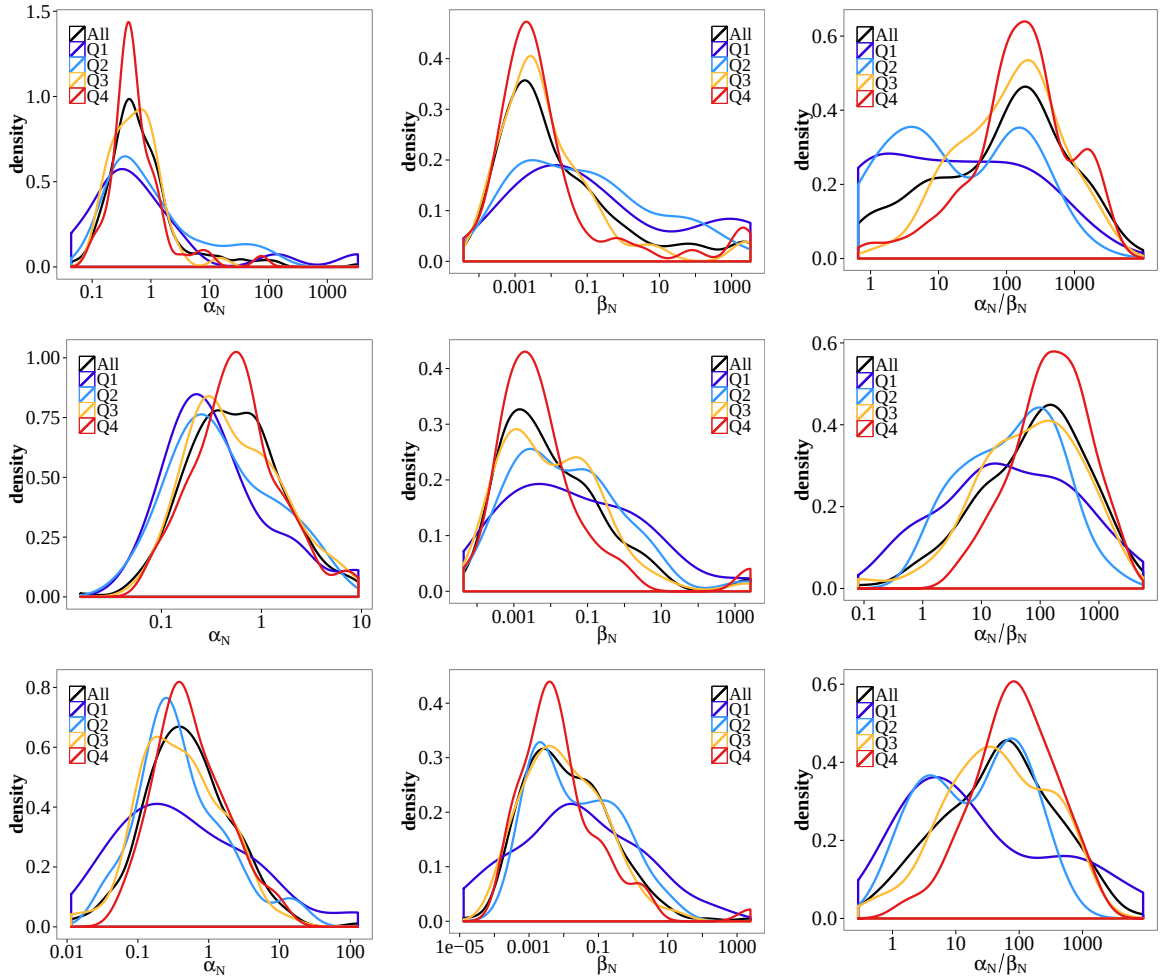

Figure 6.17: Estimated probability density of the news parameters for four different groups formed on news surprise. Q1 indicates the fist quartiles of surprise, Q2 the second and so on. Also the total distribution is plotted for comparison. First row from top shows EUR/USD, the second EUR/JPY, and the last USD/JPY

6.6.3 Relation between α_N/β_N and S

Finally, we examine the relation between the values of the parameters and the surprise indicator S defined in Section 6.2. To this end, we restricted the analysis to the news for which it was possible to calculate a surprise value, removing thus those for which the Forecast and Actual fields are not available. This left us with 214 news. Then, we divided the news and the corresponding estimated parameters in four groups based on the surprise value. The intervals are delimited by the quartiles of S. For each group, we estimate the probability distribution of the ratio α_N/β_N. The results are reported in Figure 6.17. As we have done for the impact θ, in Table 6.10 we also report the results of a $t-$test of the hypothesis that the means of the ratios α_N/β_N are equal among the group formed on surprise.

We note that for the Q1-Q4 groups the p-values are far larger than for the Q2-Q4 groups. This is a consequence of our very simple measure of surprise: despite our use of different definitions of S for indicators given in percentage, they are still penalized. This means that their surprise S falls in the lower quartiles even if in fact their impact was relevant. As a consequence, high values of the ratio α_N/β_N are found also in the Q1 group, leading to non-significant differences in the means between Q1 and Q4.

For future developments, a more sophisticated measure of surprise should be introduced, which takes into account the specificity of each type of indicator. Nevertheless, even with our simple measure S, different levels of surprise give rise to different ratio distributions. Higher ratios correspond on average to higher surprise. This is an important result. The relation between news parameters and the impact θ confirms

	groups	t	df	p-value
EUR/USD	Q1-Q4	-2.68	68.06	0.01
	Q2-Q4	-3.40	65.22	$1.1{\cdot}10^{-3}$
EUR/JPY	Q1-Q4	-0.50	25.95	0.62
	Q2-Q4	-2.96	81.70	$4.1{\cdot}10^{-4}$
USD/JPY	Q1-Q4	0.96	19.30	0.35
	Q2-Q4	-3.23	64.83	$2.0{\cdot}10^{-3}$

Table 6.10: Results of the t-test of the hypothesis that the parameters for different levels of surprise have the same mean. We checked the hypothesis of equal means both for the groups corresponding to the first and fourth quartiles of S, Q1-Q4, and for the second and fourth quartiles, Q2-Q4. The test is two sided and the Welch approximation is used for the degrees of freedom (df) to account for unequal variances in the two samples. We report the value of the t-statistic along with the computed $p-$value. The test was performed using R built-in function `t.test`.

that the model captures the impact, but was somewhat expected since θ is calculated from the same data used for the estimate. Instead, the relation between parameter and surprise is more relevant, since S is calculated from completely exogenous data and does not enter in the estimation procedure.

6.7 Influence of the news term on the endogenous parameters

In the previous sections we focused on the news term parameters α_N and β_N. Here instead we analyse how the value of the endogenous term parameters is influenced by the presence of the news term. In Figure 6.18 we compare the distributions of the endogenous parameters estimated with the model (6.4) with those estimated without the news term, i.e. with the same model of Chapter 5. The graphs are relative to USD/JPY, but the qualitative observation we make are valid also for the other pairs. For both the model with the news term and the one without it, we plot the distribution of the parameters for news with $\theta \leq \theta_{\text{median}}$ ("Low") and those with $\theta > \theta_{\text{median}}$ ("High"). We note that:

- The parameter that exhibits the most relevant difference when the news term is included is n. We note that for high impact news n takes lower values in the model that feature the news term. In fact, an important part of the activity is now explained by the news term, whereas, in the previous model, it was attributed to the endogenous term.
- The baseline intensity μ has instead higher values in the model with the news term. This is at first sight somewhat surprising, since one could expect that without the news term a higher part of the observed intensity is attributed to the baseline intensity μ. However, this effect can be due to the fact that, as we noted, n is smaller with the news term, but the news term contributes zero to the intensity before the news. Hence a higher value of μ is needed to compensate for this.
- The tail of the endogenous kernel is estimated to decay faster to zero in presence of the news term and for high impact news. Instead, for low impact news, both models find about the same values.
- Finally, the values of τ_0 are not very sensitive to the presence of the news term. Hence, it seems that it is the tail of the kernel which is more sensitive to the introduction of the news term, as τ_0 influence mainly the form of the kernel at low values of t.

The endogenous parameters are clearly influenced by the presence of the news term. In particular, after the news an important fraction of the intensity is now attributed to the news term. We note that the value of n estimated with the news term may be underestimated when no news are present (e.g. before the announcement) and this in turn leads to higher values of μ to compensate. This result is due to the fact that we employ constant parameters, while probably the "right" values of n before and after a news

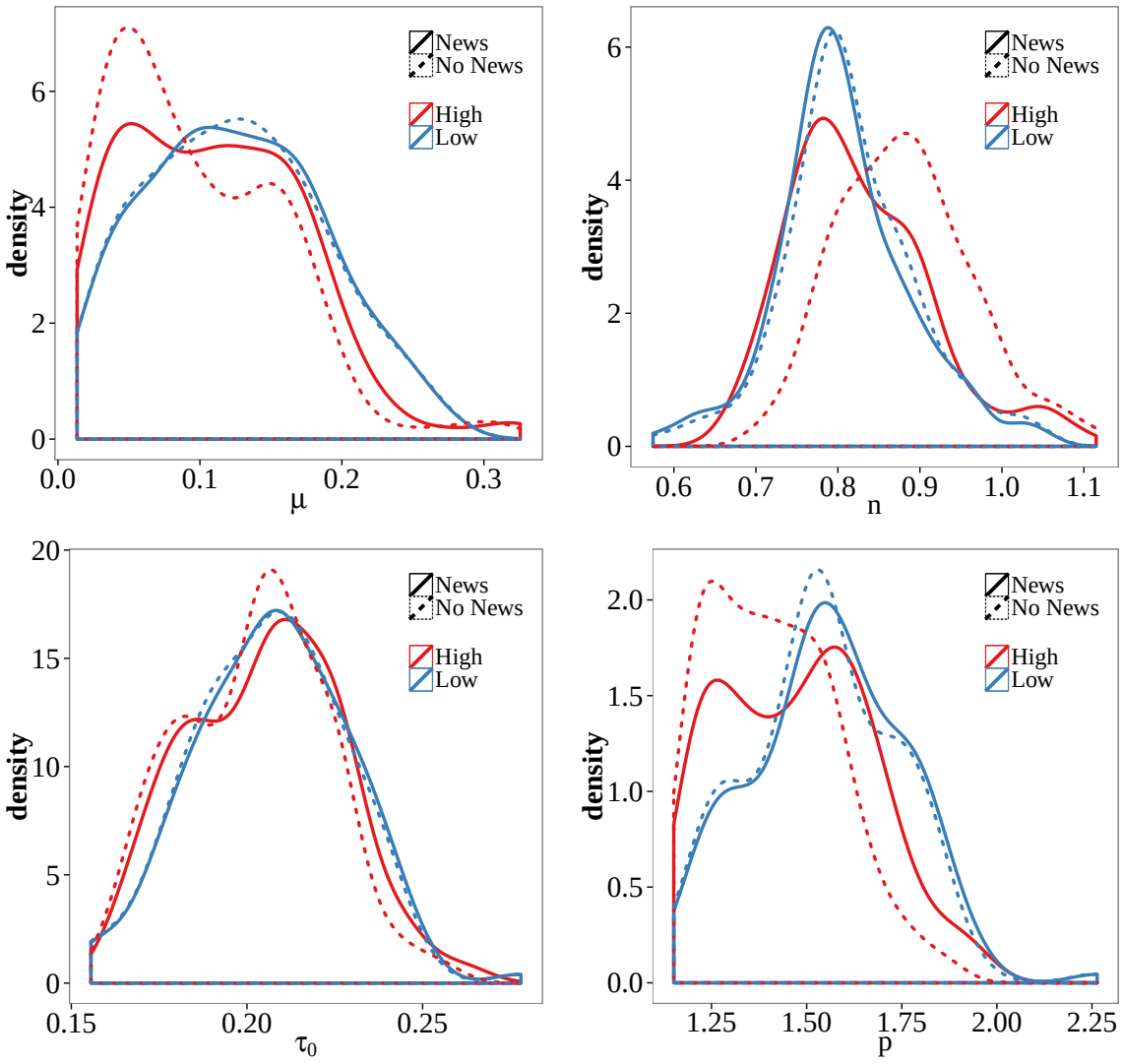

Figure 6.18: Distribution of endogenous parameters estimated on the same data with and without the news term. The graphs are for USD/JPY. "High" refers to the parameters estimated on windows where an high impact news was released, $\theta > \theta_{\text{median}}$. "Low" is for low impact news, $\theta \leq \theta_{\text{median}}$

are different. As we seen in Chapters 4 and 5 in fact, on average a large fraction of activity (≈ 0.9) is explained by the endogenous term. Our intuition is that far from news events $n \approx 1$, i.e. activity is almost entirely endogenous, whereas after an important news n becomes much smaller and the exogenous component gains weight. Future improvements should address the issue of the time dependence of the contributions.

6.8 Model selection and final remarks

In conclusion, the extension of the Hawkes framework to exogenous events we proposed appears to perform rather well. Its main strengths are the capability of the parameters α_N and β_N to capture some relevant feature of the news and the ability of the model to reproduce via simulation the observed behaviour of activity after a news release.

As a final validation tool, we used Akaike information Criterion (AIC) (Akaike, 1974) to compare the likelihoods of the model with the news term against the one without it. We remind that the AIC scores is

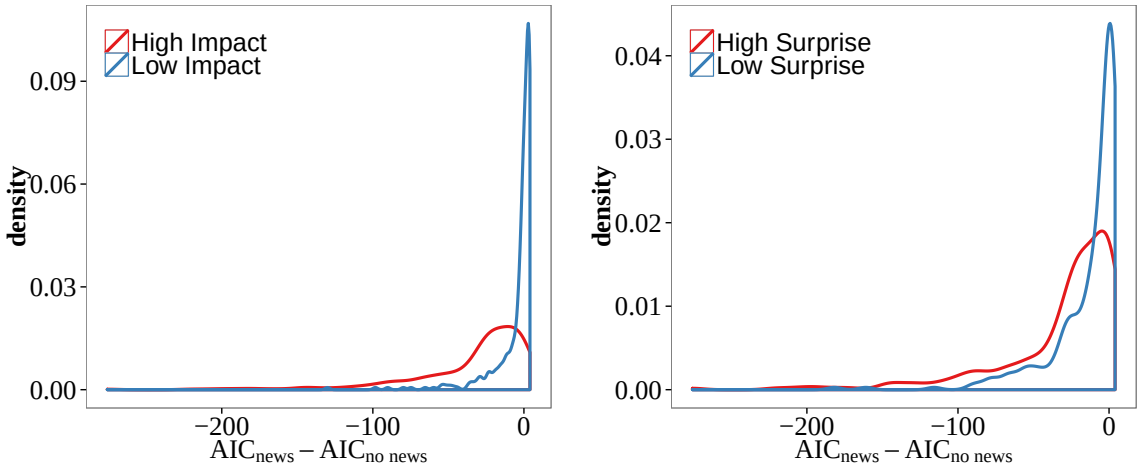

Figure 6.19: Distribution of the difference in AIC scores, $\text{AIC}_{\text{News}} - \text{AIC}_{\text{no News}}$, between the model with the news term and the one without it. Left: distributions for high impact news, $\theta > \theta_{\text{median}}$, and low impact news,$\theta \leq \theta_{\text{median}}$. Right: distributions for high surprise news $S > S_{\text{median}}$, and low surprise news, $S \leq S_{\text{median}}$. All the currency pairs are aggregated. Median values of θ are calculated before the aggregation for each pair.

defined to be

$$\text{AIC} = 2k - 2\ln\mathcal{L}$$

the model with the lowest value of AIC is the one to be preferred. Given a set of models $\{i = 1, 2, .., M\}$ the relative likelihood RL of model i with respect to the best model in the set to which corresponds $\text{AIC} = \text{AIC}_{\text{min}}$ is given by

$$\text{RL} = \exp\left(\frac{\text{AIC}_{\text{min}} - \text{AIC}_i}{2}\right). \tag{6.11}$$

First we observe that AIC scores for the models with the power law endogenous kernel are always better than those for the models with the double exponential kernel. Hence, in the following we compare the results of the models with the power law endogenous kernel that differ for the presence of the news term. In Figure 6.19 we plot the distribution of the difference in AIC scores between the model with the news term and the one without, namely $\text{AIC}_{\text{News}} - \text{AIC}_{\text{no News}}$. The distribution is calculated separately for low impact and high impact news, and for high surprise and low surprise news. The figure is obtained pooling the values from all currency pairs. Only a small fraction of the differences are positive, meaning that the model with the news term is almost always better (lower values of AIC are better). However, for low-impact and low surprise news, there is a sharp peak close to zero, meaning the model are essentially equivalent. Looking at the average relative likelihoods RL for the two groups we get:

Indicator	High	Low
θ	$5.6 \cdot 10^{-8}$	$1.0 \cdot 10^{-1}$
S	$1.5 \cdot 10^{-7}$	$1.3 \cdot 10^{-3}$

This means that for high impact news the model without the news term is about 10^{-8} times as probable as the model with the news term to minimize the information loss, i.e. the model with the news component is much better. Unsurprisingly, for low impact news the the difference between the model reduces, albeit AIC still favours the model with the news term. We mention that the QQ-plots of the residuals from the two models do not evidence such a dramatic difference as AIC scores do. This is due to the fact that the comparison of the residuals

$$\Lambda_i = \int_{t_i}^{t_{i+1}} \lambda(t)\mathrm{d}t$$

is sensitive to the overall value of $\lambda(t)$ in each interval $[t_i, t_{i+1}]$. Hence if the two models both find similar values of λ, as observed, then the residuals have also the same values. The model with the news term gives almost the same value of $\lambda(t)$ as the model without the exogenous term, it is the decomposition in endogenous and exogenous contributions which makes the difference. This emerges clearly form AIC and also from simulations such as the one presented in Figure 6.8.

A final note: an important limitation of our model is that, as it was presented, it treats equally all the news. So if more than one news is present in the estimation interval and the news have different characteristics that lead to very different impacts, the model is expected to average the differences or to be influenced by the most relevant ones. Future developments can address this issue for example by introducing a surprise-dependent amplitude $\alpha_N = \alpha_N(S)$. Limiting the dependence on surprise to the amplitude allows to conserve the recursion relation that speeds up likelihood computation. However, if the number of exogenous events is limited, one can also let β_N depending on surprise.

Chapter 7

Expanding the information set: Hawkes processes with kernel dependent on price changes

In this Chapter, we present an extension of the Hawkes model seen so far that aims at including in the kernel more information than the sole series of past times at which events took place. All the material contained in this Chapter is original. The results we present are intended to be preliminary.

So far, we used only the time information of our datasets. In the expression of conditional intensity function $\lambda(t|\mathcal{F}_t)$ only the previous events of the process appear, i.e. the intensity was conditioned only on process *history*, $\mathcal{F}_t = \mathcal{H}_t = \{t_0, t_1, \dots, t_i\}$, with $t_i \leq t$. However, order book data contain much more information (price, volume, etc.) that can be exploited to enhance the model. This means using a larger filtration, $\mathcal{F}_t = \{(t_0, Z_0), (t_1, Z_1), \dots, (t_i, Z_i)\} \supset \mathcal{H}_t$, where Z_i indicates other relevant information. In a Hawkes process framework, this further information, Z_i, can be introduced by making the parameters Z_i-dependent (Bowsher, 2007). For example, in the simplest case of a single exponential kernel, we can set:

$$\lambda(t|\mathcal{F}_t) = \mu + \int \alpha(Z_i) e^{-\beta(Z_i)(t-s)} \mathrm{d}N_s$$

Note that the functional dependence of the parameters on Z_i has to be specified parametrically. Here we present an extension of the Hawkes models of Chapter 5 that includes the information on how much the mid price has changed at time t_i.

7.1 The model

The model develops from the observation that it is not realistic that all quote changes will influence the future arrival rate in the same way, regardless of the magnitude of the correspondent price changes. It is instead reasonable to expect that large price moves trigger more activity than small ones.

We modify our Hawkes model for FOREX quote changes in accordance with this observation by using a double exponential kernel, with amplitudes depending on the magnitude of mid quote changes. Hence, now we make the αs depend on the absolute changes Δ in the mid quote, measured in ticks. In principle also the β could depend on Δ, however, keeping the βs constant greatly simplifies computations, in that the recursive formula for the likelihood is still applicable. The model thus reads

$$\lambda(t|\mathcal{F}_t) = \mu + \sum_{t_i<t} \alpha_A(\Delta_i) e^{-\beta_A(t-t_i)} + \sum_{t_i<t} \alpha_B(\Delta_i) e^{-\beta_B(t-t_i)} \tag{7.1}$$

As we mentioned, it is necessary to specify parametrically how the αs depend on Δ. We choose a linear form for $\alpha(\Delta)$:

$$\alpha_p(\Delta) = \alpha_p^0 + \alpha_p^1 \cdot \Delta, \text{ with } p = A, B \tag{7.2}$$

The idea is that a larger change in Δ causes higher future activity thanks to a larger amplitude α, that we set proportional to Δ itself. In this way, we have now a total of 7 parameters: μ, α_A^0, α_A^1, β_A, α_B^0, α_B^1, β_B. Again, as in Chapter 5, the A subscript denotes the largest of the βs (shortest time scale).

Let us point out that in developing the model we present here, we considered the process Δ as given and independent of the activity process we model. In particular, we assume that the Δs are independent identically distributed random variables and that the process governing Δ is independent of the quotes counting process. In other words, we are assuming that the next value of Δ is randomly extracted with a certain probability distribution that does not depend on how much time has passed since the last event, nor on the previous value of Δ.

We choose to test this extension with the double exponential kernel because here the two time scales are completely separated and this allows to see which time scale is more influenced by the information on the magnitude of the mid price increase. For the log-likelihood, from equation (2.32) we have

$$\ln \mathcal{L} = -\int_0^T \lambda(t|\mathcal{F}_t)\mathrm{d}t + \sum_{t_i<t} \ln \lambda(t|\mathcal{F}_t) \tag{7.3}$$

(A) (B)

where, denoting with N the number of events in $[0,T]$ and defining $t_{n+1} = T$

$$\begin{aligned}
\text{(A)} &= \mu T + \sum_{p=A,B} \int_0^T \sum_{t_i<t} \alpha_p(\Delta_i) e^{-\beta_p(t-t_i)}\mathrm{d}t = \\
&= \mu T + \sum_{p=A,B} \sum_{i=1}^{N} \int_{t_i}^{t_{i+1}} \sum_{t_i<t} \alpha_p(\Delta_i) e^{-\beta_p(t-t_i)}\mathrm{d}t = \\
&= \mu T + \sum_{p=A,B} \left\{ \frac{\alpha_p(\Delta_1)}{\beta_p} \left[1 - e^{-\beta_p(t_2-t_1)}\right] + \frac{\alpha_p(\Delta_1)}{\beta_p} \left[e^{-\beta_p(t_2-t_1)} - e^{-\beta_p(t_3-t_1)}\right] + \right. \\
&\left. + \frac{\alpha_p(\Delta_2)}{\beta_p} \left[1 - e^{-\beta_p(t_3-t_2)}\right] + \dots \right\} = \\
&= \mu T + \sum_{p=A,B} \sum_{t_i<t} \frac{\alpha_p(\Delta_i)}{\beta_p} \left[1 - e^{-\beta_p(T-t_i)}\right].
\end{aligned} \tag{7.4}$$

For (B), we introduce the quantity $R_p(i)$

$$\begin{aligned}
R_p(i) &= \sum_{t_j<t_i} \alpha_p(\Delta_j) e^{-\beta_p(t_i-t_j)} = \\
&= \alpha_p(\Delta_{i-1}) e^{-\beta_p(t_i-t_{i-1})} + \sum_{t_j<t_{i-1}} \alpha_p(\Delta_j) e^{-\beta_p(t_i-t_j+t_{i-1}-t_{i-1})} = \\
&= e^{-\beta_p(t_i-t_{i-1})} \left[\alpha(\Delta_{i-1}) + R_p(i-1)\right]
\end{aligned} \tag{7.5}$$

with $R_p(1) = 0$. Writing R_p in this form allows to compute it recursively, reducing time computation cost of the log-likelihood from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$. Hence

$$\text{(B)} = \sum_{t_i} \ln \left(\mu + \sum_{p=A,B} R_p(i) \right) \tag{7.6}$$

Ultimately, we have for the log-likelihood

$$\ln \mathcal{L} = -\mu T - \sum_{p=A,B} \sum_{t_i<t} \frac{\alpha_p(\Delta_i)}{\beta_p} \left[1 - e^{-\beta_p(T-t_i)}\right] + \sum_{t_i} \ln \left(\mu + \sum_{p=A,B} R_p(i) \right). \tag{7.7}$$

Finally, let us examine the condition for stationarity, we must have $n = \int_0^\infty w(t)\mathrm{d}t < 1$. In this case, since $\alpha = \alpha(\Delta)$, we have to integrate also over the distribution of Δ. Then, if we indicate with $P(\Delta)$ the probability density of Δ, the condition is written

$$\sum_i P(\Delta_i) \int_0^\infty \alpha_A(\Delta_i) e^{-\beta_A(t)} + \alpha_B(\Delta_i) e^{-\beta_B(t)} < 1. \tag{7.8}$$

Note that the actual distribution $P(\Delta)$ is not known, so the empirical distribution has to be used in testing for stationarity.

7.2 Empirical test

We fit model (7.1) on the same FOREX data of Chapter 5. We estimate the best parameters day by day by using maximum likelihood. The optimization algorithm *nlminb* was used as in Chapter 5. Our FOREX dataset presents an interesting feature for the test of this model. In fact, as was already pointed out, on 23 September 2012 the tick size was increased by a factor of five. Since now the model includes the information on Δ, we expect to see some difference after the tick increase.

Figures 7.1, 7.2, and 7.3 show our results for the amplitudes α. The transition from the small-tick regime to the large-tick one is clearly visible. Moreover, the influence on the dynamics of the magnitude Δ of the mid change is more pronounced at the short time scale, at least in the small-tick period. Table 7.1 reports the typical uncertainties associated to the parameter estimates.

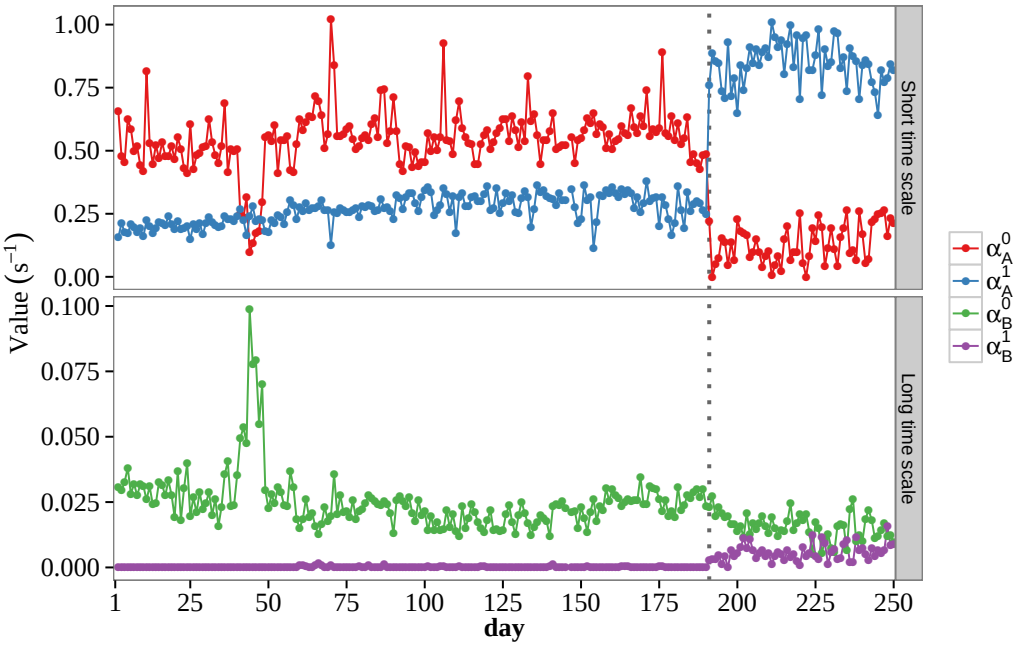

Figure 7.1: Time series of the amplitude estimates for EUR/USD. On day 191 (23 September 2012), EBS increased the tick size by a factor of five on all the currencies pairs we are studying.

During the small tick period, α_B^1 is found to be zero (within the uncertainty) for all days. Each change in the best quotes has the same effect on the intensity at long time scale ($\beta_B^{-1} \approx 15 - 30\,s$), regardless of the magnitude of the change. At the short time scale ($\beta_A^{-1} \approx 0.3 - 0.7\,s$) instead, the contribution to the amplitude α_A from the term proportional to Δ is the most relevant.

In fact, in Table 7.2 the average contributions to the total amplitude are reported, together with $\langle\Delta\rangle$ which is the average value of the increments Δ. In Figure 7.4 the time series of the ratio $(\alpha_A^1 \cdot \langle\Delta\rangle)/\alpha_A^0$ is plotted. Here $\langle\Delta\rangle$ is calculated for each day. The term $\alpha_A^1 \cdot \langle\Delta\rangle$ is always dominant (except for EUR/JPY

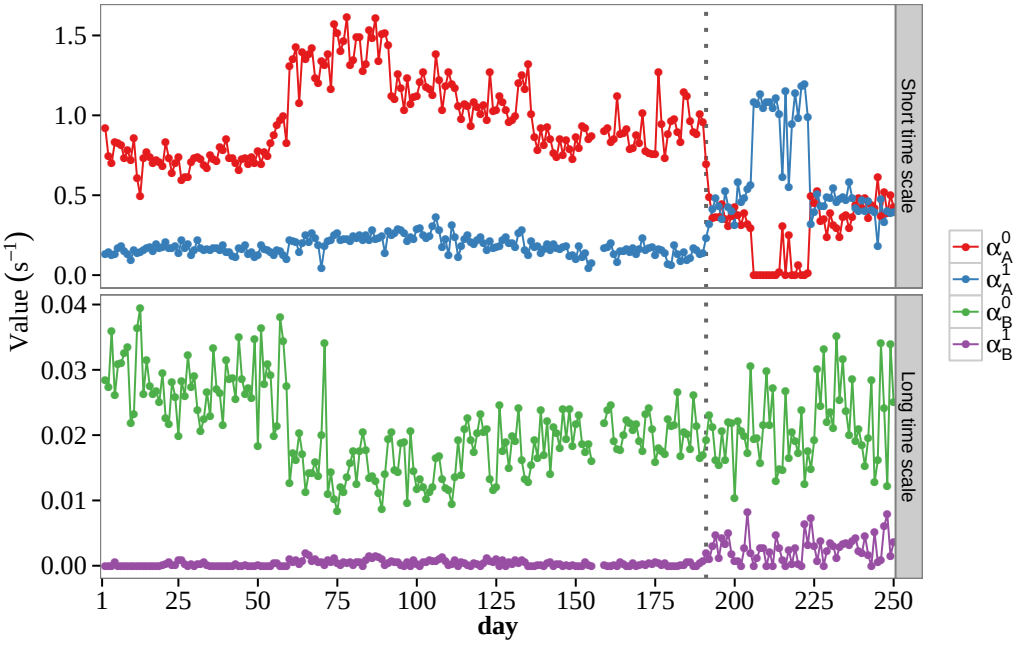

Figure 7.2: Time series of the amplitude estimates for EUR/JPY. On day 191 (23 September 2012), EBS increased the tick size by a factor of five on all the currencies pairs we are studying.

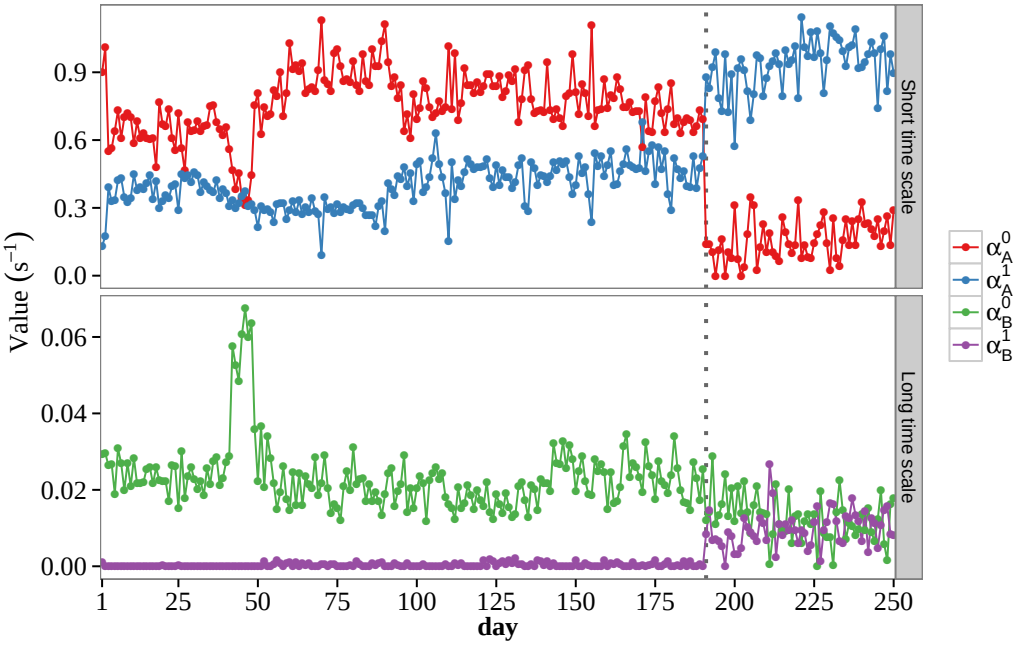

Figure 7.3: Time series of the amplitude estimates for USD/JPY. On day 191 (23 September 2012), EBS increased the tick size by a factor of five on all the currencies pairs we are studying.

	μ	α_A^0	α_A^1	β_A	α_B^0	α_B^1	β_B
Small tick	$5-10\%$	$4-5\%$	$2-4\%$	$1-2\%$	$10-15\%$	$>100\%$	$8-10\%$
Large tick	$5-10\%$	$40-60\%$	$5-10\%$	$2-4\%$	$20-30\%$	$40-50\%$	$8-10\%$

Table 7.1: Order of magnitude of the uncertainties on the parameter estimates calculated from the inverse of the Hessian at the maximum. Values are in percentage.

in the small tick regime), but its role becomes even more relevant after the tick increase. After the tick increase, Δ is almost always equal to one tick. However, when an increase of more than one tick is recorded, it has a much larger effect. Moreover, the effect in the large tick regime is noticeable even on the long time scale, albeit here the dominant term remains α_B^0 which does not depend on Δ.

Small tick

	$\langle\Delta\rangle$	$\alpha_A^0\ (s^{-1})$	$\alpha_A^1 \cdot \langle\Delta\rangle\ (s^{-1})$	$\alpha_B^0\ (s^{-1})$	$\alpha_B^1 \cdot \langle\Delta\rangle\ (s^{-1})$
EUR/USD	3.2	0.56 (35)	0.83 (23)	0.025 (44)	$3 \cdot 10^{-4}$ (270)
EUR/JPY	3.9	0.99 (28)	0.70 (30)	0.021 (16)	$3 \cdot 10^{-4}$ (122)
USD/JPY	2.5	0.76 (18)	0.98 (24)	0.023 (37)	$3 \cdot 10^{-4}$ (153)

Large tick

	$\langle\Delta\rangle$	$\alpha_A^0\ (s^{-1})$	$\alpha_A^1 \cdot \langle\Delta\rangle\ (s^{-1})$	$\alpha_B^0\ (s^{-1})$	$\alpha_B^1 \cdot \langle\Delta\rangle\ (s^{-1})$
EUR/USD	1.28	0.13 (57)	1.08 (10)	0.016 (31)	0.0073 (57)
EUR/JPY	1.31	0.29 (65)	0.80 (48)	0.021 (27)	0.0034 (80)
USD/JPY	1.18	0.16 (57)	1.10 (12)	0.013 (47)	0.0114 (24)

Table 7.2: Average values of the increase Δ of the mid quote (in ticks), and averages of the amplitudes α_p^0 and $\alpha_p^1 \cdot \langle\Delta\rangle$, before and after the tick increase. Standard deviations are reported in parentheses (values are in percentage).

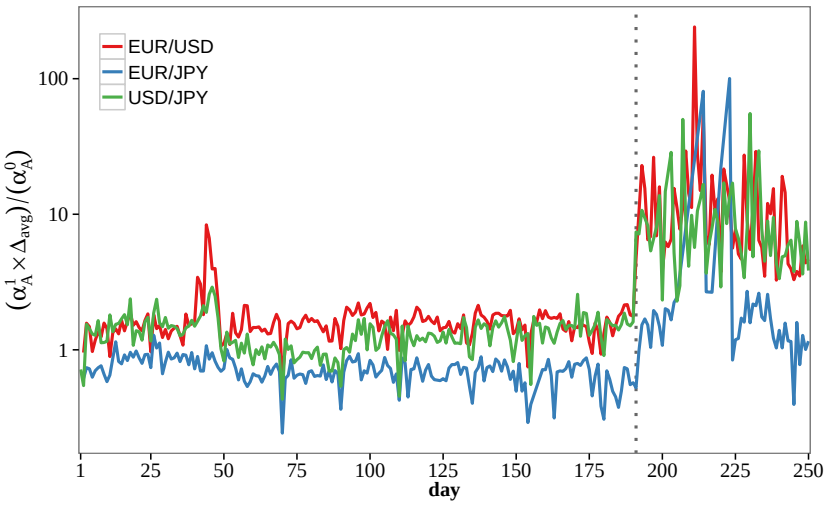

Figure 7.4: Time series of the ratio $(\alpha_A^1 \cdot \langle\Delta\rangle)/\alpha_A^0$ for the three currency pairs.

Days 42-48 of both EUR/USD and USD/JPY present a peak in α_B^0 and a corresponding drop in α_A^0. We compare in Figure 7.5 the duration distribution in these days with the one in the previous days. The days with the anomalous behaviour feature a higher number of durations around 1 s. This in turn causes higher values of α_B^0, leading to the observed peak. Since the peak is not detected on EUR/JPY, it seems that it is

connected to USD. However, the reason of the higher-than-normal number of durations around 1 s is not known to us.

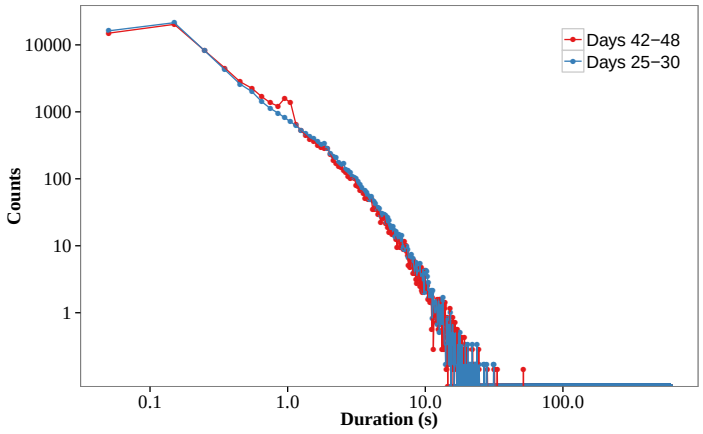

Figure 7.5: Duration distribution of days 42-48 compared with the one of days 25-30. EUR/USD sample. For each group of days, the points represent the average number of durations in the corresponding bin. The bin size is 0.1 s, and the abscissa of the points are the mid points of the bins. The scale of the graph is log-log. A clear difference between the distributions is seen around 1 s.

We note also that between day 210 and day 225 the behaviour of the α in EUR/JPY is quite strange. The reason for this result is not clear however, since activity and distribution of Δ are analogous to the other days of the large-tick period.

With regard to the stationarity condition (7.8), for the model presented, the branching ratio n has a simple expression:

$$\begin{aligned} n &= \sum_i P(\Delta_i) \int_0^\infty \alpha_A(\Delta_i) e^{-\beta_A(t)} + \alpha_B(\Delta_i) e^{-\beta_B(t)} = \\ &= \sum_i P(\Delta_i) \left\{ \frac{\alpha_A^0}{\beta_A} + \frac{\alpha_A^1 \Delta_i}{\beta_A} + \frac{\alpha_B^0}{\beta_B} + \frac{\alpha_B^1 \Delta_i}{\beta_B} \right\} = \\ &= \frac{\alpha_A^0}{\beta_A} + \frac{\alpha_B^0}{\beta_B} + \left(\frac{\alpha_A^1}{\beta_A} + \frac{\alpha_B^1}{\beta_B} \right) \langle \Delta \rangle. \end{aligned} \tag{7.9}$$

Figure 7.6 shows the values of n obtained for the three pairs. The parameter n is calculated day by day using the average value of Δ in each day. The values are almost equal to those obtained with the basic model of Chapter 5. In fact, as we shall see in the next section the values of α and β are essentially equal in the two models. The model appears to capture some relevant features of the data. First of all, the magnitude Δ of the mid change is found to have an effect on market activity. This effect is clear immediately after the event, and tends to fade away as the time distance to the event increases. Finally, the effect increases in the large tick size regime. All these findings seem very reasonable and thus support the validity of this model extension.

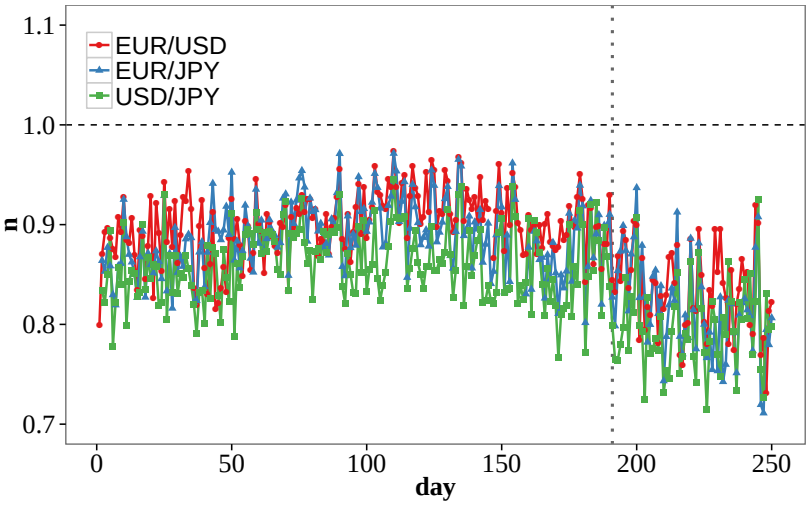

Figure 7.6: Time series of the branching ratio n for the three currency pairs. The dashed line indicates the critical value $n = 1$, while the dotted line signals separates the small tick size period and the large tick size one.

7.3 Comparison with the original model

Now we examine how the enhanced model (7.1) compare with the basic double exponential model of Chapter 5. Figure 7.7 reports the QQ-plot of the residuals against the exponential distribution for the model (7.1) and the one tested in Chapter 5 that does not take into account Δ. The difference in the residual distribution between the two model is almost unnoticeable. As noted in Chapter 6 however, this only means that the values of $\lambda(t)$ calculated in the two models are pretty similar on average. This is confirmed by looking at Table 7.3 , where we compare the values of α and β obtained in the two models. We calculate day by day the ratio between corresponding parameters in the two models. It turns out that they match each other quite closely, all the ratios are very close to one and compatible with one within the uncertainty. This confirms the claim that the value of $\lambda(t)$ are on average very close one to each other in the two models, and explains the result observed in the QQ-plot.

(a) Small tick

(b) Large tick

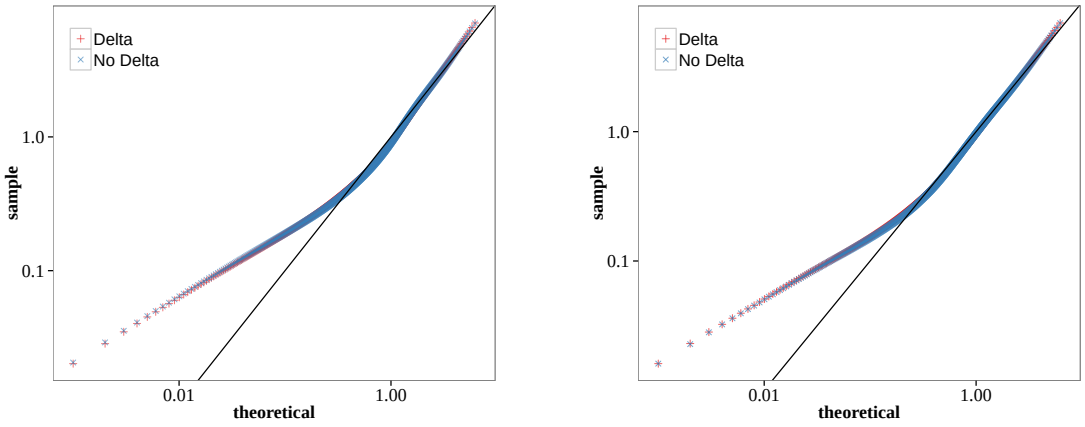

Figure 7.7: Comparison of the models with and without conditioning the α parameters on Δ. QQ-plot of the residuals against exponential distribution, before and after the tick size increase. Scales are logarithmic.

However, the values of the likelihood are quite different between the two models. We use Akaike Information Criterion as in Chapter 6 to compare the likelihood accounting for the different number of parameters. In Figure 7.8 we plot the distribution of the difference in AIC scores between the models,

	$\frac{\alpha_A^0+\alpha_A^1\cdot\langle\Delta\rangle}{\alpha_A^{\text{plain}}}$	$\frac{\beta_A^\Delta}{\beta_A^{\text{plain}}}$	$\frac{\alpha_B^0+\alpha_B^1\cdot\langle\Delta\rangle}{\alpha_B^{\text{plain}}}$	$\frac{\beta_B^\Delta}{\beta_B^{\text{plain}}}$
EUR/USD	0.98 ± 0.04	1.02 ± 0.06	1.3 ± 0.4	1.3 ± 0.3
EUR/JPY	1.0 ± 0.4	1.02 ± 0.06	1.1 ± 0.1	1.11 ± 0.06
USD/JPY	0.977 ± 0.012	0.99 ± 0.03	1.11 ± 0.12	1.11 ± 0.10

Table 7.3: Ratios between corresponding parameters in the two models: with and without conditioning on Δ. The values are calculated day by day, and the average values are shown. The uncertainty is the standard deviation.

aggregating all currency pairs. A lower AIC scores indicates a better performance. The model with the Δ dependence has lower AIC in all cases. The different values for the small-tick and large-tick regimes are due to the lower sample size in the large-tick period. In fact the relative difference in AIC scores, $(\text{AIC}_\Delta - \text{AIC}_{\text{plain}})/\text{AIC}_{\text{plain}}$ is almost constant and in fact a bit larger in the large tick regime.

In any case, the relative likelihood of the plain model to the "delta" model

$$\text{RL} = \exp(\text{AIC}_\Delta - \text{AIC}_{\text{plain}})$$

is essentially zero, meaning that AIC clearly prefers the enhanced model.

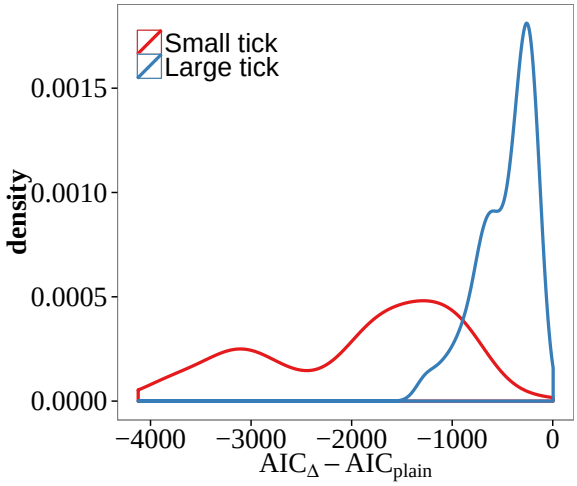

Figure 7.8: Distribution of the AIC scores differences between the two models. The results for the small tick regime and for the large tick one are plotted separately. All currency pairs are aggregated.

The results we presented in this Chapter are still preliminary. The work has to be completed with the analysis of simulated paths generated with the model, and an analogous extension could be implemented also with the power law kernel.

As a final remark, despite we considered the process Δ as given and independent of the activity process we model, this is not the real case, since is very plausible that the increments are not independent to each other and that their magnitude can be influenced by the activity process itself. To solve this issue one has to consider a *marked* point process that describes both the time occurrence of the events and their associated marks (Δ in this case).

Conclusion

We now summarize our main results and outline some possible future developments. In this thesis we used Hawkes processes to model market activity. Hawkes processes have proven to be able to capture many features of the real data and to reproduce pretty well the observed dynamics. Moreover, they provide a quantitative measure of endogenous and exogenous contributions to market activity.

In Chapter 4 and 5 we discussed the capabilities of a univariate Hawkes model to fit both equity and FOREX data. In accordance to recent results published in the literature, we found that markets appears to operate near criticality ($n \approx 1$), supporting the idea that most of the market activity is endogenously generated. We also confirmed that a power law kernel delivers a better fit, while a single exponential specification is completely inadequate. We also showed how certain peculiarities of the FOREX dataset, first of all the coarse time resolution with which order book data are sampled, can condition the model performance.

The content of Chapter 6 and 7 represents our main contribution. In Chapter 6 we developed a Hawkes model that allows external shocks to influence the dynamics together with the characteristic self-exciting component. This, to the best of our knowledge, has not been done before. The results of the empirical test of the model showed, quite interestingly, that the benefices in terms of goodness-of-fit measured by the QQ-plot are hardly noticeable. However, looking at the likelihood or, even more significantly, at how well a simulated path reproduces the observed data leaves no doubt on the superior description obtained when one includes the "news" term. This has an important implication on the debate over the origin of market activity, namely that a model without a proper external-shock term can overestimate the amount of endogenously-generated activity in proximity of the external event. Besides, this can remain unnoticed since, as we said, goodness of fit can appear equally good without including the news term. We found that, although the endogenous contribution remains dominant, in the case of an important external shock, the exogenous contribution becomes relevant. Then, time-dependent parameters that can account for time varying endogenous and exogenous contributions are probably needed. In this way, one could reproduce an almost entirely self-exciting market away from main external events, while accounting for non negligible exogenous contributions in proximity of external shocks, which seems the observed market behaviour.

Hawkes processes appear to be a promising tool for the development of a faithful model of financial markets dynamics at high frequency. In particular, they offer themselves to various possible extensions that can provide increasingly more realistic models. We demonstrated in Chapter 6 the possibility of including other types of events. Parameters can also be made dependent on other information available in the data. For example, we suggested that the news parameters of our model can be made dependent on the surprise brought by the news, i.e. the difference between the news and its ex ante expectation, in order to allow the contribution of different news to be weighted appropriately.

In Chapter 7 we developed the idea of making parameters dependent on further information. While this possibility is known in the literature, we are not aware of any practical implementation and empirical test of a model like the one we presented. We illustrated a model that takes into account the magnitude of the mid-quote changes in the description of the dynamics, the intuition being that large jumps in the mid price trigger more activity than small ones do. We showed that the model is indeed capable of capturing some relevant features of the data. The influence of the price change magnitude is found to be relevant

mainly at a short time distance. Moreover, also the transition to a larger tick size, as the one happened on EBS platform, is clearly detected by the model. Once again, standard goodness-of-fit test employed in the Hawkes literature do not highlight any performance improvement with respect to the basic model, while likelihood-based model selection clearly indicates the superiority of the new model.

Finally, future work towards a complete market model should aim at combining the time-dimension description - how events arrange in time - with the price dynamics itself - how much price change. This could be potentially achieved developing marked Hawkes process models, in which *when* the price move is strictly interconnected with *how much* it moves.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on 19*(6), 716–723.

Ardia, D., K. M. Mullen, B. G. Peterson, and J. Ulrich (2012). *DEoptim: Differential Evolution in R*. version 2.2-1.

Bacry, E., K. Dayri, and J. Muzy (2012). Non-parametric kernel estimation for symmetric Hawkes processes. application to high frequency financial data. *The European Physical Journal B-Condensed Matter and Complex Systems 85*(5), 1–12.

Bacry, E., S. Delattre, M. Hoffmann, and J.-F. Muzy (2013). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance 13*(1), 65–77.

Bauwens, L. and P. Giot (1997). The logarithmic ACD model: an application to market microstructure and NASDAQ. *CORE Discussion Papers 1997089, Université catholique de Louvain*.

Bauwens, L. and N. Hautsch (2009). Modelling financial high frequency data using point processes. In T. Mikosch, J.-P. Kreiß, R. A. Davis, and T. G. Andersen (Eds.), *Handbook of Financial Time Series*, pp. 953–979. Springer Berlin Heidelberg.

Bormetti, G., L. M. Calcagnile, M. Treccani, F. Corsi, S. Marmi, and F. Lillo (2013). Modelling systemic cojumps with hawkes factor models. *arXiv preprint arXiv:1301.6141*.

Bouchaud, J.-P. (2011). The endogenous dynamics of markets: price impact, feedback loops and instabilities. *Lessons from the Credit Crisis. Risk Publications*.

Bouchaud, J.-P. and M. Potters (2003). *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge University Press.

Bowsher, C. G. (2002). Modelling security market events in continuous time: Intensity based, multivariate point process models. Working Paper. `http://www.nuff.ox.ac.uk/economics/papers/2002/w22/NuffWP_final.pdf`.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics 141*(2), 876 – 912.

Brémaud, P. and L. Massoulié (2001). Hawkes branching point processes without ancestors. *Journal of applied probability 38*(1), 122–135.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 2. The new algorithm. *IMA Journal of Applied Mathematics 6*(3), 222–231.

Carstensen, L. (2010). *Hawkes processes and combinatorial transcriptional regulation*. Dissertation, University of Copenhagen.

Chakraborti, A., I. M. Toke, M. Patriarca, and F. Abergel (2009). Econophysics: Empirical facts and agent-based models. *arXiv preprint arXiv:0909.1974*.

Cont, R. and P. Tankov (2004). *Financial Modelling With Jump Processes*. Chapman and Hall/CRC, London.

Daley, D. J. and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Springer, New York.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica 66*, 1127–1162.

Filimonov, V. and D. Sornette (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E 85*(5), 056108.

Fletcher, R. (1970). A new approach to variable metric algorithms. *The computer journal 13*(3), 317–322.

Gay, D. M. (1990). *Computing Science Technical Report No. 153: Usage Summary for Selected Optimization Routines*. Murray Hill, NJ 07974: AT& T Bell Laboratories.

Gillemot, L., J. D. Farmer, and F. Lillo (2006). There's more to volatility than volume. *Quantitative Finance 6*(5), 371–384.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of computation 24*(109), 23–26.

Hardiman, S. J., N. Bercot, and J.-P. Bouchaud (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *arXiv preprint arXiv:1302.1405*.

Harris, T. E. (2002). *The theory of branching processes*. Courier Dover Publications.

Hasbrouck, J. (2007). *Empirical market microstructure: The institutions, economics and econometrics of securities trading*. Oxford University Press New York.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika 58*, 83–90.

Hawkes, A. G. and D. Oakes (1974). A Cluster Process Representation of a Self-Exciting Process. *Journal of Applied Probability 11*(3), 493–503.

Hewlett, P. (2006). Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes, Ecole Polytechnique*, pp. 6–8.

Ivanov, P. C., A. Yuen, B. Podobnik, and Y. Lee (2004). Common scaling patterns in intertrade times of u. s. stocks. *Physical Review E 69*, 056107.

Joulin, A., A. Lefevre, D. Grunberg, and J. Bouchaud (2008). Stock price jumps: News and volume play a minor role. *Wilmott Magazine*, 1–7.

King, M. R., L. Osler, and D. Rime (2011). Foreign exchange market structure, players and evolution. Technical report, Norges Bank.

LeRoy, S. F. (2006). Excess volatility. *The New Palgrave Dictionary of Economics, 2nd Edition. Palgrave Macmillan 13*.

LeRoy, S. F. and R. D. Porter (1981). The present-value relation: Tests based on implied variance bounds. *Econometrica 49*(3), 555–74.

Lewis, P. A. and G. S. Shedler (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly 26*(3), 403–413.

Malkiel, B. G. (1999). *A random walk down Wall Street: Including a life-cycle guide to personal investing*. WW Norton & Company.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business 36*(4), 394–419.

Mantegna, R. N. and H. E. Stanley (1999). *Introduction to econophysics: correlations and complexity in finance*. Cambridge University Press.

Mike, S. and J. Farmer (2008). An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control 32*(1), 200–234.

Nash, J. C. and R. Varadhan (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software 43*(9), 1–14.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics 30*(1), 243–261.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory 27*(1), 23–31.

Plerou, V., P. Gopikrishnan, L. A. Nunes Amaral, X. Gabaix, and H. Eugene Stanley (2000). Economic fluctuations and anomalous diffusion. *Physical Review E 62*(3), R3023–R3026.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rasmussen, J. G. (2011). Temporal point processes: the conditional intensity function. Lecture notes. `http://people.math.aau.dk/~jgr/teaching/punktproc11/tpp.pdf`.

Revolution Analytics (2012). *foreach: Foreach looping construct for R*. R package version 1.4.0.

Roberts, H. (1967). Statistical versus clinical prediction in the stock market. University of Chicago, Unpublished manuscript.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial management review 6*(2), 41–49.

Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of computation 24*(111), 647–656.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *The American Economic Review 71*(3), 421–436.

Sornette, D. (2006). Endogenous versus exogenous origins of crises. In *Extreme events in nature and society*, pp. 95–119. Springer.

Soros, G. (2009). *The crash of 2008 and what it means: the new paradigm for financial markets*. PublicAffairs.

Toke, I. M. (2011a). An introduction to Hawkes processes with applications to finance. Course slides. `http://fiquant.mas.ecp.fr/ioane_files/HawkesCourseSlides.pdf`.

Toke, I. M. (2011b). "Market Making" in an Order Book Model and Its Impact on the Spread. In F. Abergel, B. K. Chakrabarti, A. Chakraborti, and M. Mitra (Eds.), *Econophysics of Order-driven Markets*, New Economic Windows, pp. 49–64. Springer Milan.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.

Acknowledgements

I would like to express my deepest gratitude to Prof. Fabrizio Lillo for his guidance in all the time of research and writing of this thesis. He helped me continuously and with an exceptional patience.

I am grateful to the HSBC eFX trading group headed by Mr. Richard Anthony. They provided us with the FOREX data and they were very kind to me and helpful when I visited them in London. A special thanks goes to Dr. Paris Pennesi for the inspiring conversation and precious advices.

I wish also to thank Prof. Riccardo Mannella for reading the manuscript and helping improving it.

Mamma, Papà and Meggy, you know how much I love you and how much I owe to you. Lucia, thank you for your love and your invaluable support.